Development and application of non-linear image enhancement and multi-sensor fusion techniques for hazy and dark imaging conditions NCC-1-01030

Summary of Research
submitted to
NASA Langley Research Center
Hampton, Virginia 23681

Principal Investigator
Zia-ur Rahman
Research Associate Professor

Department of Applied Science
College of William & Mary
PO Box 8795
Williamsburg, VA 23187-8795
ATTN: Grants & Research Administration

February 15, 2005

This is the final report for NASA cooperative agreement NCC-1-01030 and covers the period from 01 March, 2001 to 28 February, 2004. Enclosed are copies of the conference papers and journal articles for which the research was at least partially supported by the grant.

1 Description of Research

The purpose of this research was to develop enhancement and multi-sensor fusion algorithms and techniques to make it safer for the pilot to fly in what would normally be considered Instrument Flight Rules (IFR) conditions, where pilot visibility is severely restricted due to fog, haze or other weather phenomenon. We proposed to use the non-linear Multiscale Retinex[1, 2, 3, 4] (MSR) as the basic driver for developing an integrated enhancement and fusion engine.

When we started this research, the MSR was being applied primarily to grayscale imagery such as medical images, or to three-band color imagery, such as that produced in consumer photography: it was not, however, being applied to other imagery such as that produced by infrared image sources. Even for such imagery, the MSR processed images, because of the inherent sharpness present in the computation, showed a discernible reduction in haziness. Additionally, because of the dynamic range compression and lightness constancy properties, the MSR processing resulted in an almost complete elimination of the problems due to low brightness, provided of course that the low brightness level did not result in initial loss of data when the image was acquired. Figure 1 shows examples of these situations.

However, we felt that it was possible by using the MSR algorithm in conjunction with multiple imaging modalities such as long-wave infrared (LWIR), short-wave infrared (SWIR), and visible spectrum (VIS), we could substantially improve over the then state-of-the-art enhancement algorithms, especially in poor visibility conditions. We proposed the following tasks:

1. Investigate the effects of applying the MSR to LWIR and SWIR images. This consisted of optimizing the algorithm in terms of surround scales, and weights for these spectral bands.
2. Fusing the LWIR and SWIR images with the VIS images using the MSR framework to determine the best possible representation of the desired features.
3. Evaluating different mixes of LWIR, SWIR and VIS bands for maximum fog and haze reduction, and low light level compensation.
4. Modifying the existing algorithms to work with video sequences.

Over the course of the 3 year research period, we were able to accomplish these tasks and report on them at various internal presentations at NASA Langley Research Center, and in presentations and publications elsewhere. A description of the work performed under the tasks is provided in Section 2. The complete list of relevant publications during the research periods is provided in Section 5. This research also resulted in the generation of intellectual property (Section 5.1)

Figure 1: Left row: low brightness level image and the MSR processed image: the effects of low brightness levels have been eliminated with respect to the signal-to-noise ratio of the acquisition device. Right row: original effects of haze caused by fog have been reduced by MSR processing.

referenced by LaRC Case Number LAR 16570-1. Though a patent has not yet been applied for, NASA Langley has determined that this technology is of use and have retained the rights for future filings.

2 Tasks

2.1 Multi-sensors

Investigate the effects of applying the MSR to LWIR and SWIR images. Optimize the algorithm in terms of surround scales, and weights.

Multi-sensor data acquired during test flights on the 757-Aries platform were used to examine the issue of weights and scales for multi-sensors. The multi-sensor data in the present case consists of data from three sensors: Long-wave infrared (LWIR), short-wave infrared (SWIR), and visible band (VIS). Each sensor provides a different view of the same scene though with different spatial resolution, field-of-view (FOV), and spectral sensitivity. Table 2.1 shows the characteristics of the various sensors[5]. The main reason for using a sensor suite, rather than a single sensor, is because features that are apparent in one modality, may or may not be apparent in another: the SWIR image may contain features that are not apparent in the VIS band and vice versa; similarly for the LWIR and VIS, and LWIR and SWIR bands.

	LWIR	**SWIR**	**VIS**
Pixel Resolution (nominal)	$320H \times 240V$	$320H \times 240V$	$542H \times 497V$ (RGB)
Optics FOV	$39°H \times 29°V$	$34°H \times 25°V$	$34°H \times 25°V$
Detector readout frame rate	60Hz	60Hz (typical)	30Hz (interlaced)

2.1.1 LWIR

The LWIR sensor provides the capability to see features that are at a temperature difference from their surroundings even through imaging conditions that would usually render a VIS camera useless. For example, an automobile would be clearly visible against a road through clouds or haze that would make a VIS image useless. The LWIR images tend to have poor contrast resolution because the image is really a depiction of the temperature variation in the scene. If there is not a significant degree of temperature variating between objects in a scene, as is the case when there is fog present, then the overall images also have less detail. Since the MSR enhances small variations in a region with respect to its surround, the overall impact of the MSR is to provide an image with greater contrast resolution and detail, Figure 2 (top-row). The default settings with which the MSR was used for grayscale and consumer photography were insufficient for the task of enhancing details in the LWIR image. By experimenting with the imagery that was acquired during several flights tests, we were able to adjust the MSR parameters, primarily contrast and brightness controls, that

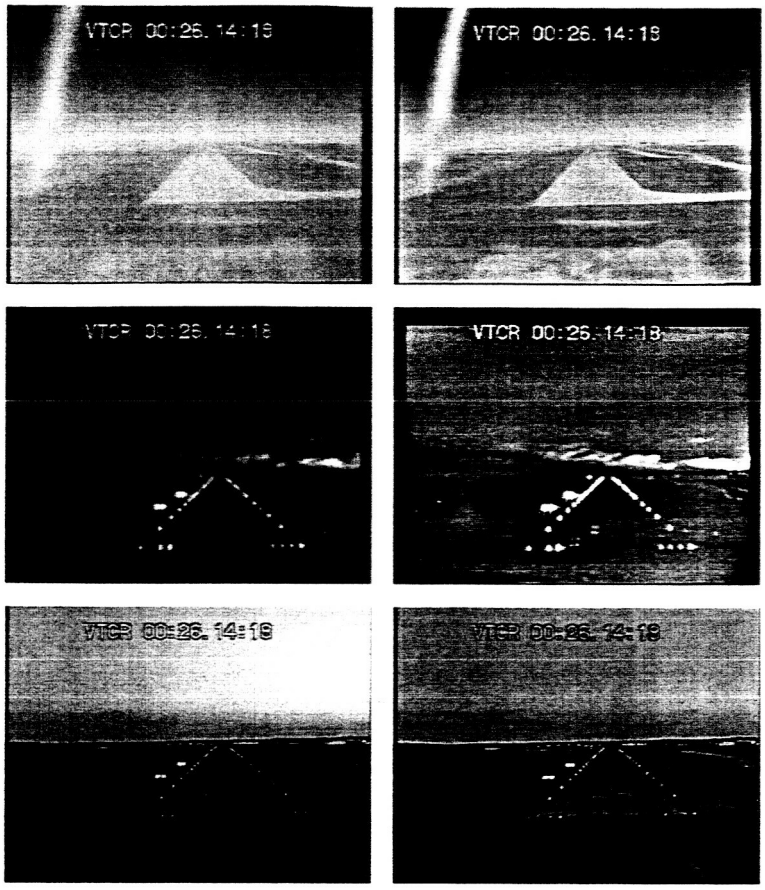


Figure 2: The impact of MSR processing on imagery: top-row, LWIR; middle-row, SWIR; bottom-row, VIS. The original (left) image has poor contrast making it difficult to discern features of interest. The MSR processed image (right) enhances the features of interest with respect to their surroundings making them considerably more discernible.

gave us a much better automatic enhancement. We should note, however, that the process can be controlled manual should different contrast and brightness levels are requires. We should also note, that for over 90% of the images, such adjustment does not need to be made.

2.1.2 SWIR

The SWIR imagery combines a bit of LWIR imagery with VIS imagery. However, it is very evident from the data that the SWIR imagery has very poor dynamic range in normal visibility conditions. We should note, though, that the settings used for this sensor were optimized for detecting runway lights in night-time lighting conditions. These settings made the sensor worthless for detecting any other features. Because of the poor dynamic range these settings produced, the resulting images also have very poor contrast. Application of the MSR increases the dynamic range and contrast

of the SWIR imagery. Figure 2 (middle-row) shows the impact of the MSR on SWIR imagery: several small features that are essentially buried in the original SWIR image become very visible in the processed data. As in the case of LWIR, the default MSR processing was insufficient for enhancing the SWIR data stream to the desired level so we had to optimize MSR parameters for this operation by using the flight data.

2.1.3 VIS

Though the visible images are traditionally not thought to be of much use in poor visibility conditions, we have developed image enhancement schemes that greatly increase their useful even in dimly lit environments. This is discussed in more detail in Section 3. Figure 2 (bottom-row) shows the impact of MSR processing on a VIS image. Again, much like in the case of LWIR and SWIR, the MSR processed image shows details that is not discernible in the original image. Since we had previously optimized the MSR for used with VIS imagery, we did not have to much optimization to get desired results. We did note that there was much more information available in the visible band, even under poor lighting and weather conditions, than we had previously conceived. In fact, because of the fact that we are used to experiencing VIS imagery, there is no *learning curve* that one has to master in order to interpret this data stream. So, if processed properly, the VIS data may be as, if not more, useful as the other two sources. We do, however, present a *fused* data stream to the pilot, so the individual benefits and eccentricities of the different data streams get combined into a single data stream that contains the salient information from all the data streams. More on this in Section 2.2.

2.2 Fusion

Fusing the LWIR and SWIR images with the VIS images using the MSR framework to determine the best possible representation of the desired features.

Image fusion provides the mechanism to present the information that is available in the three sensor modalities into a single image that contains all the salient features from the different modalities. There are a number of different techniques to perform fusion. One can do simple arithmetic averaging over the three modalities, or one can merge feature based upon a maximum response, or something else altogether[6]. There is, however, one (major) underlying assumption for all of these techniques: the images from the different sensors are rectified and co-registered. By rectification we mean that regardless of the original FOVs and spatial resolutions, the rectified images all have the same instantaneous FOV (IFOV), i.e., are mapped with respect to a common grid. By co-registration we mean that the major features in the different images lie on the same rectified grid location.

Co-registration of image features can be relatively straightforward, or tricky, depending upon the characteristics of the sensors. As Table 2.1 shows clearly, the three sensor differ in spatial

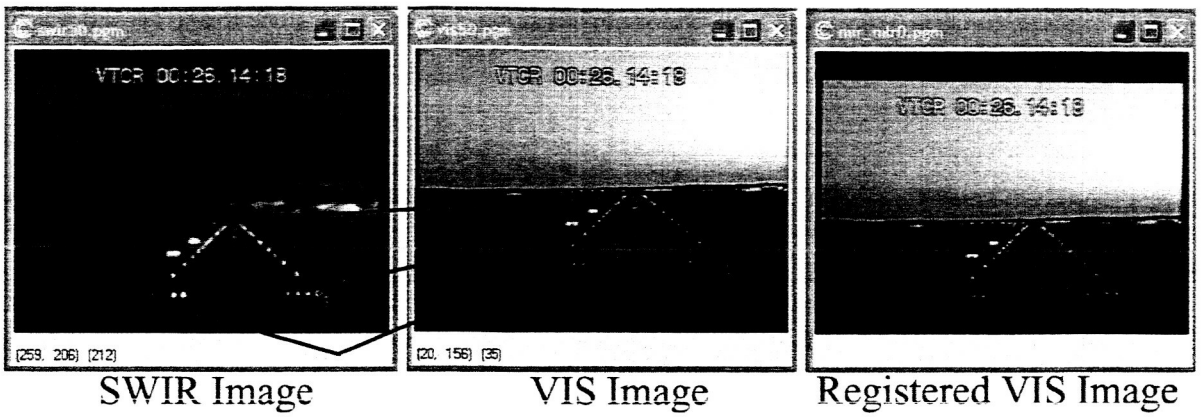


Figure 3: Co-registering VIS and SWIR bands.

resolution and FOV. This means that IFOV is different for each sensor. The first order of business then is to remap the data from each sensor so that their IFOVs are the same. There are several constraints on how rectification is performed in the context of image fusion:

1. Smallest FOV determines the FOV of the fused components. Hence the data with the larger FOV needs to be cropped so that the FOVs of all the data are the same.

2. The coarsest IFOV determines the direction of adjustment. In other words, if a sensor possesses a finer IFOV than another sensor, its IFOV will have to coarsened by resampling in order to accommodate the larger grid size of the poorer sensor.

The procedure for rectification is given in detail in [7].

If the sensors are perfectly bore-sighted, i.e., object from different sensors form on the same rectified grid point, then, disregarding the slight parallax difference because the sensors are not co-located, rectifying the data automatically registers the data. If, however, the sensors are *not* bore-sighted, additional data transforms are required to register the data. Unfortunately, in the case of the multi-sensor pod that was used in flight tests, it is very evident from the acquired data, see Figure 2, that aside from the differences in physical characteristics, the cameras are also not bore-sighted. There are significant vertical and horizontal offsets, as well as some rotation and shear differences. All of these can be corrected for by applying affine transforms to the data. However, this is not an easy problem for automation. We developed a set of tools that can be used to register the data manually by using control points in the various data streams. Figure 3 shows an example of this technique. Once a set of transforms has been obtained, it can be applied to each frame of the data stream because the relative alignment of the sensors remains constant over a flight test. The material is presented in Hines, et al[7], a copy of which is attached.

Once the data has been co-registered, as was said above, various different techniques can be used to fuse the data. We addressed this problem in Rahman et al[6]. The novel part of the approach presented in this paper was the use of the MSR as an integrated fusion and enhancement engine.

Each single data stream was used as a single band in the MSR process so each band was processed using a different surround scale and a different weighting function. The impact of changing the surround scale is to create output imagery in which different features are enhanced. For instance small surround scales produce enhancement of fine details with relative loss of color, and large surround scales preserve lightness and color at the cost of details[1, 4]. Further, since the contrast, brightness and sharpness, can be controlled by appropriate gain, and offset, we were able to convert the post-registration enhancement and fusion into a one-step operation where the MSR is used to enhance and fuse the data. Figure 4 shows an example of this fusion.

Additionally, it should be noted that different image applications and imaging conditions can lead to a different set of requirements on which images have the more important features. Though the MSR is used to bring out salient features in all the data streams, one can control in what amount the data mixtures are fused so as to give one data type priority over another. This was also investigated in [6], where we looked at assigning different weights to the three different sensor outputs and examining the fused data for best results.

We will be developing these techniques further in current and future research. Copies of relevant publications accompany this report.

3 Intelligent Enhancement

As mentioned earlier, the parameters for the image are sensor dependent. This means that we have to pre-set parameters for each image stream which may lead to sub-optimal results. We have developed a new method that combines autonomous decision making with a suite of image enhancement and sharpening algorithms and adds a level of autonomy and intelligence to image enhancement and processing. This method is called the visual servo (VS) and is described in NASA LaRC disclosure # LAR 16570-1. A copy of the disclosure has not been attached but is available from the NASA LaRC patent attorney's office. The VS derives its intelligence by examining brightness, contrast and sharpness characteristics of each image[8, 9, 1]. Based upon these computed values, the VS classifies the image as being of GOOD contrast and sharpness, GOOD contrast, but POOR sharpness, and so on. Figure 5 shown an example of these classifications. Based upon each combination of classifications, the images pass through a series of processing steps that modify their contrast, brightness and sharpness until certain exit criteria are fulfilled. Figure 6 shows a block diagram that describes this process.

Figure 7 gives an example of this process for the VIS sensor. In addition to showing the servo output, this figure also shows that with proper processing—servo for example—images taken in low light levels can become quite useful.

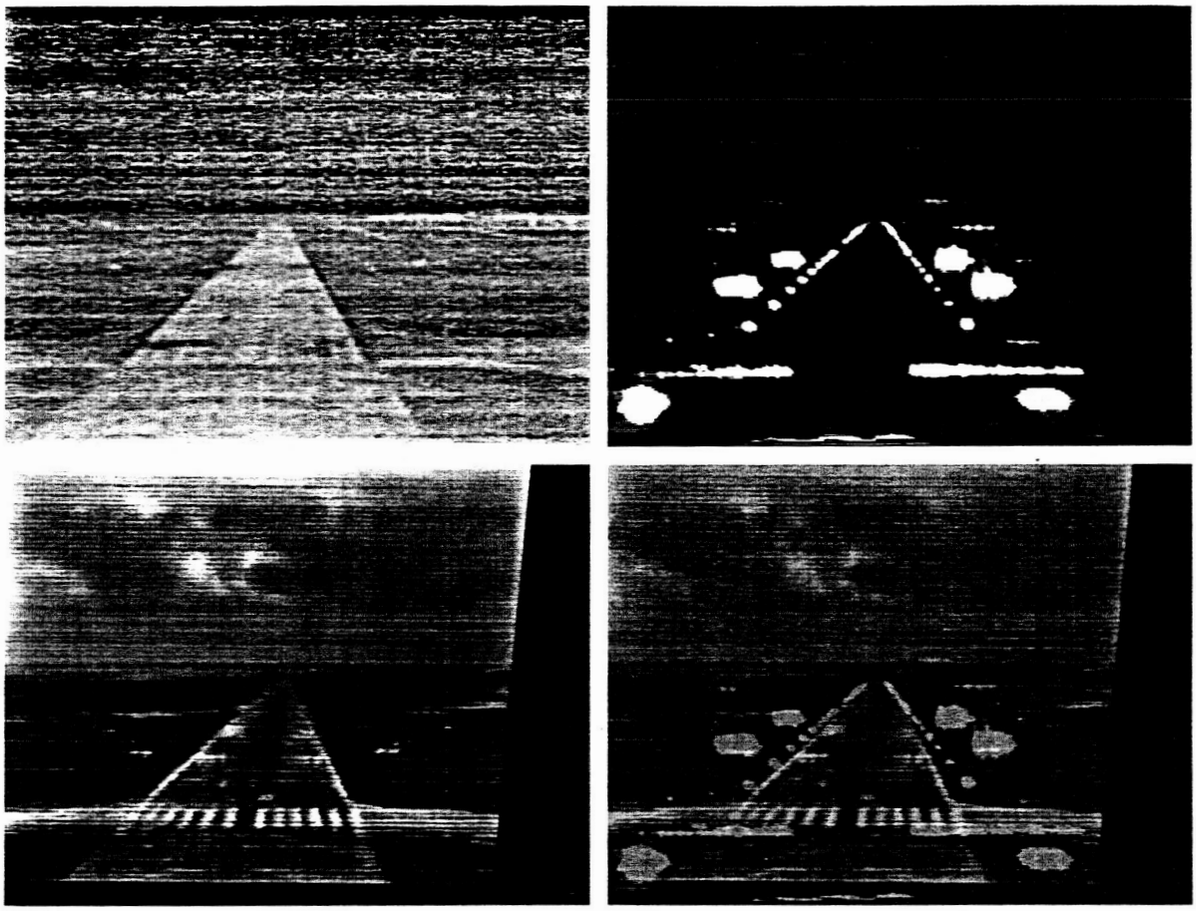


Figure 4: LWIR (top-left), SWIR (top-right), RGB (bottom-left), and fused (bottom-right). The fused image synergistically combines the information present in the LWIR, SWIR, and RGB data streams into a single representation. These data streams have been enhanced prior to fusion.

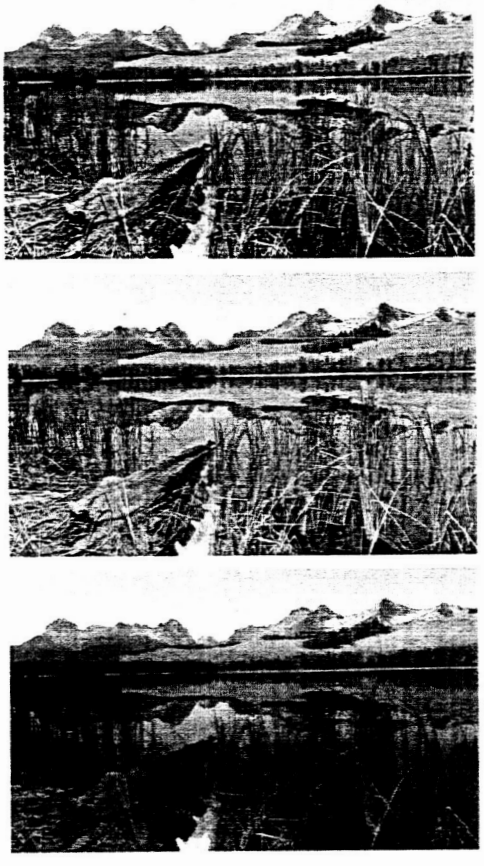


Figure 5: (a) Visual Measures for automating visual assessment: images are assigned to one of three global classes, {Excellent, Good, Poor}. The classes are bases on global and regional brightness and contrast measures.

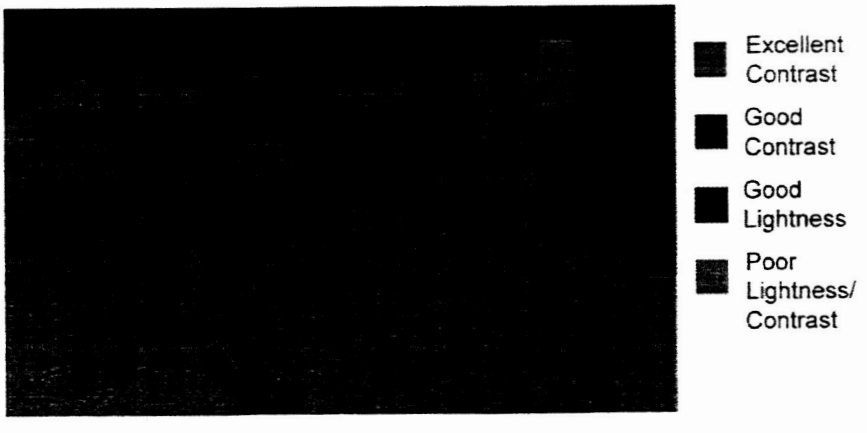


Figure 5: (b) Visual map showing regional classes: the combination of the regional classes is used to derive the global classification.

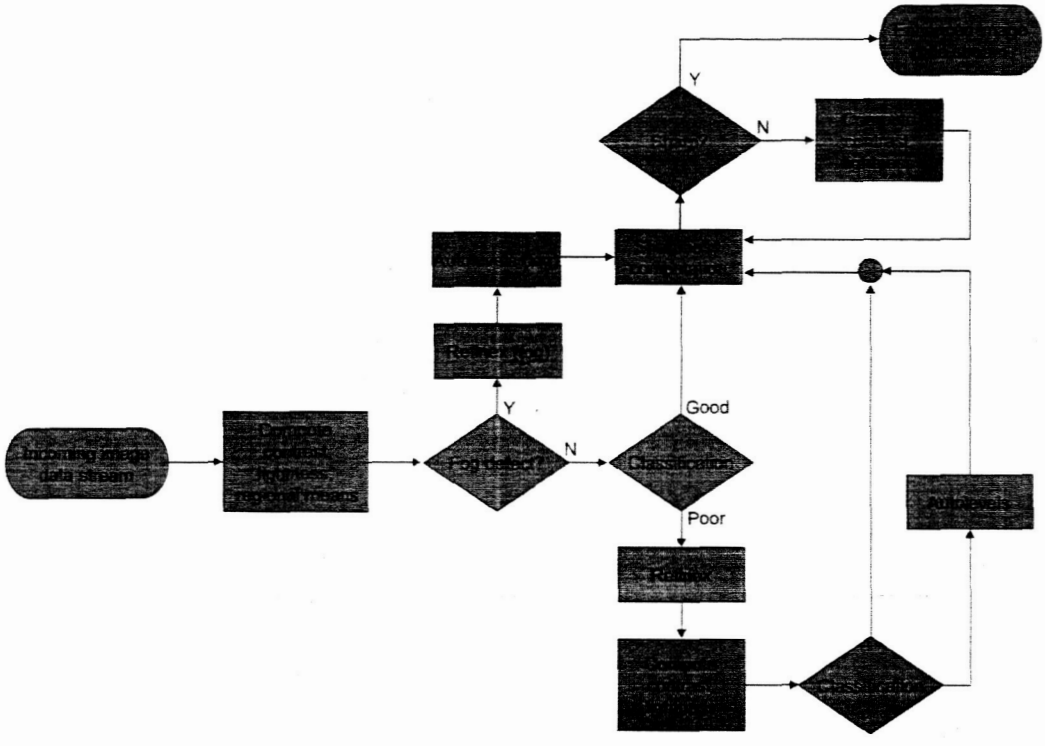


Fig. 2. Servo system

Figure 6: Block diagram showing the servo mechanism.

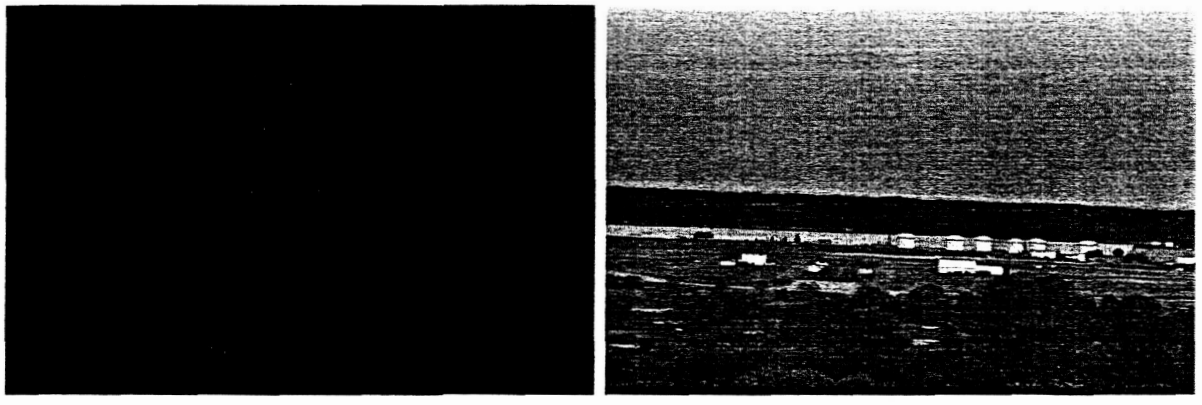


Figure 7: Original image (left) taken under low light level conditions can be virtually useless. The servo (right) automatically set the parameters for the MSR process and additional processing steps to produce an image that contains useful information.

4 Video processing

During the period when this research was performed, we did not have the capability to perform real-time video processing. Since then we have been successful in implementing a single scale version of the MSR that operates on monochrome images[10, 11]. We did, however, develop software procedures to perform MSR enhancement on video streams:

1. Decompose the video stream into individual video frames.
2. Process these individual video frames with the MSR
3. Synthesize the processed individual frames into a video stream.

While straightforward in concept, several interesting issues arose when we performed this processing. The MSR produces an output based upon image characteristics—whether there are sharp shadows, whether the lighting conditions are varying, etc. When used with video frames, we found that the automatic mode of processing was resulting in visible "flashes" in the processed imagery. These "flashes" seemed to occur in frames that were temporally close, where the MSR was compensating for changes in light levels by automatically adjusting brightness and contrast parameters. Hence we had to modify the video processing scheme to make the adjustment of parameters static for the whole video stream. Though this results in sub-optimal per frame performance, the overall enhanced video is much more useful.

5 Publications

Publications are listed chronologically in each category.

5.1 Intellectual Property

1. "A Computational Visual Servo: Automatic Measurement and Control for Smart Image Enhancement", D. J. Jobson. Z. Rahman, G. A. Woodell, LaRC patent disclosure #LAR 16570-1.

   Summary:
   Based upon the scientific insights gained from Retinex image processing (US patent 5,991,456, pending US 09/88,701, and 09/888,816, Australian patent 713076) and the statistical convergence of a large set of enhanced images, a set of absolute visual measures (VM) was developed which consists of contrast, lightness, and sharpness measures for arbitrary digital images. These measures, in turn, serve as a rudimentary form of visual intelligence which provides the foundation for an automatic servo control system for image enhancement. Based upon the VM scores and classifications, enhancement modules that include Retinex image enhancement, autolevels, and sharpening are invoked. For any digital image, the servo will

cycle among the modules until the image is enhanced to acceptable VM scores, or until the allowable degrees of enhancement have been exhausted. Further, as a special case of interest to aviation safety, the servo contains a unique module for detecting turbid imaging conditions (fog/smoke/haze) and applies a unique combination of enhancement processing to optimize this class of image enhancement for maximum scene clarity.

5.2 Journal Article

1. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex Processing for Automatic Image Enhancement", Journal of Electronic Imaging, Volume 13, Number 1, January 2004.

Abstract:
There has been a revivification of interest in the Retinex computation in the last six or seven years, especially in its use for image enhancement. In his last published concept (1986) for a Retinex computation, Edwin Land introduced a center/surround spatial form which was inspired by the receptive field structures of neurophysiology. With this as our starting point we developed the Retinex concept into a full scale automatic image enhancement algorithm—the Multi-Scale Retinex with Color Restoration (MSRCR)—which combined color constancy with local contrast/lightness enhancement to transform digital images into renditions that approach the realism of direct scene observation. Recently we have been exploring the fundamental scientific questions raised by this form of image processing: namely, (i) is the linear representation of digital images adequate in visual terms in capturing the wide scene dynamic range? (ii) can visual quality measures using the MSRCR be developed? and (iii) is there a canonical, i.e., statistically ideal, visual image? The answers to these questions can serve as the basis for automating visual assessment schemes, which, in turn, are a primitive first step in bringing visual intelligence to computers.

5.3 Conference Publications

1. Z. Rahman and D. J. Jobson, "Information Theoretic Analysis of Noise Sources in Image Formation," Visual Information Processing XII, Proc. SPIE 5108, (2003).

Abstract:
When a linear image acquisition device captures an image, several noise sources impact the quality of the final (digital) image. We have previously examined the impact of these noise sources in terms of their impact on the total amount of information that is contained in the image, and in terms of their impact on the restorability of the image data. In this paper, we will examine the effect of each of the noise sources on final image quality.

2. G. D. Hines, Z. Rahman, D. J. Jobson and G. A. Woodell, "Multisensor Image Registration For An Enhanced Vision System," Visual Information Processing XII, Proc. SPIE 5108, (2003).

Abstract:
An Enhanced Vision System (EVS) utilizing multi-sensor image fusion is currently under development at the NASA Langley Research Center. The EVS will provide enhanced images of the flight environment to assist pilots in poor visibility conditions. Multi-spectral images obtained from a short wave infrared (SWIR), a long wave infrared (LWIR), and a color visible band CCD camera, are enhanced and fused using the Retinex algorithm. The images from the different sensors do not have a uniform data structure: the three sensors not only operate at different wavelengths, but they also have different spatial resolutions, optical fields of view (FOV), and bore- sighting inaccuracies. Thus, in order to perform image fusion, the images must first be co-registered. Image registration is the task of aligning images taken at different times, from different sensors, or from different viewpoints, so that all corresponding points in the images match. In this paper, we present two methods for registering multiple multi-spectral images. The first method performs registration using sensor specifications to match the FOVs and resolutions directly through image resampling. In the second method, registration is obtained through geometric correction based on a spatial transformation defined by user selected control points and regression analysis.

3. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Feature visibility limits in the non-linear enhancement of turbid images," Visual Information Processing XII, Proc. SPIE 5108, (2003)

Abstract:
The current X-ray systems used by airport security personnel for the detection of contraband, and objects such as knives and guns that can impact the security of a flight, have limited effect because of the limited display quality of the X-ray images. Since the displayed images do not possess optimal contrast and sharpness, it is possible for the security personnel to miss potentially hazardous objects. This problem is also common to other disciplines such as medical Xrays, and can be mitigated, to a large extent, by the use of state-of-the-art image processing techniques to enhance the contrast and sharpness of the displayed image. The NASA Langley Research Centers Visual Information Processing Group has developed an image enhancement technology that has direct applications to this problem of inadequate display quality. Airport security X-ray imaging systems would benefit considerably by using this novel technology, making the task of the personnel who have to interpret the X-ray images considerably easier, faster, and more reliable. This improvement would translate into more accurate screening as well as minimizing the screening time delays to airline passengers. This technology, Retinex, has been optimized for consumer applications but has been applied to medical X-rays on a very preliminary basis. The resultant technology could be incorporated into a new breed of commercial x-ray imaging systems which would be transparent to the screener yet allow them to see subtle detail much more easily, reducing the amount of time needed for screening while greatly increasing the effectiveness of contraband detection and thus public safety.

4. D. J. Jobson, Z. Rahman, and G. A. Woodell, "The statistics of visual representation," Visual

Information Processing XI, Proc. SPIE 4736, (2002) (Invited paper)

Abstract:
The experience of retinex image processing has prompted us to reconsider fundamental aspects of imaging and image processing. Foremost is the idea that a good visual representation requires a non-linear transformation of the recorded (approximately linear) image data. Further, this transformation appears to converge on a specific distribution. Here we investigate the connection between numerical and visual phenomena. Specifically the questions explored are: (1) Is there a well-defined consistent statistical character associated with good visual representations? (2) Does there exist an ideal visual image? And (3) what are its statistical properties?

5. Z. Rahman, D. J. Jobson, G. A. Woodell, and G. D. Hines, "Multi-sensor fusion and enhancement using the Retinex image enhancement algorithm ," Visual Information Processing XI, Proc. SPIE 4736, (2002)

   Abstract:
   A new approach to sensor fusion and enhancement is presented. The retinex image enhancement algorithm is used to jointly enhance and fuse data from long wave infrared, short wave infrared and visible wavelength sensors. This joint optimization results in fused data which contains more information than any of the individual data streams. This is especially true in turbid weather conditions, where the long wave infrared sensor would conventionally be the only source of usable information. However, the retinex algorithm can be used to pull out the details from the other data streams as well, resulting in greater overall information. The fusion uses the multiscale nature of the algorithm to both enhance and weight the contributions of the different data streams forming a single output data stream.

6. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Retinex processing for automatic image enhancement," Human Vision and Electronic Imaging VII, SPIE Symposium on Electronic Imaging, Proc. SPIE 4662, (2002)

   Abstract:
   In the last published concept (1986) for a Retinex computation, Edwin Land introduced a center/surround spatial form, which was inspired by the receptive field structures of neurophysiology. With this as our starting point we have over the years developed this concept into a full scale automatic image enhancement algorithm—the Multi-Scale Retinex with Color Restoration (MSRCR) which combines color constancy with local contrast/lightness enhancement to transform digital images into renditions that approach the realism of direct scene observation. The MSRCR algorithm has proven to be quite general purpose, and very resilient to common forms of image pre-processing such as reasonable ranges of gamma and contrast stretch transformations. More recently we have been exploring the fundamental scientific implications of this form of image processing, namely: (i) the visual inadequacy of the linear representation of digital images, (ii) the existence of a canonical or statistical ideal

visual image, and (iii) new measures of visual quality based upon these insights derived from our extensive experience with MSRCR enhanced images. The lattermost serves as the basis for future schemes for automating visual assessment—a primitive first step in bringing visual intelligence to computers.

7. D. J. Jobson, Z. Rahman, and G. A. Woodell, "The Spatial Aspect of Color and Scientific Implications of Retinex Image Processing," SPIE International Symposium on AeroSense, Proceedings of the Conference on Visual Information Processing X, April 2001.

Abstract:
The history of the spatial aspect of color perception is reviewed in order to lay a foundation for the discussion of retinex image processing. While retinex computations were originally conceived as a model for color constancy in human vision, the impact on local contrast and lightness is even more pronounced than the compensation for changes in the spectral distribution of scene illuminants. In the multiscale retinex with color restoration (MSRCR), the goal of the computation is fidelity to the direct observation of scenes. The primary visual shortcoming of the recorded image is that dark zones such as shadow zones are perceived with much lower contrast and lightness than for the direct viewing of scenes. Extensive development and testing of the MSRCR led us to form several hypotheses about imaging which appear to be basic and general in nature. These are: (1) the linear representation of the image is not usually a good visual representation, (2) retinex image enhancements tend to approach a statistical ideal which suggests the existence of a canonical "visual image", and (3) the mathematical form of the MSRCR suggests a deterministic definition of visual information which is the log of the spectral and spatial context ratios for any given image. These ideas imply that the imaging process should be thought of, not as a replication process whose goal is minimal distortion, but rather as a profound non-linear transformation process whose goal is a statistical ideal visual representation. These insights suggest new directions for practical advances in bringing higher levels of visual intelligence to the world of computing.

References

[1] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," *Journal of Electronic Imaging* **13**(1), pp. 100–110, 2004.

[2] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," in *Human Vision and Electronic Imaging VII*, B. E. Rogowitz and T. N. Pappas, eds., pp. 390–401, Proc. SPIE 4662, 2002.

[3] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing* **6**, pp. 451–462, March 1997.

[4] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965–976, July 1997.

[5] C. L. Tiana, J. R. Kerr, and S. D. Harrah, "Multispectral uncooled infrared enhanced-vision system for flight test," in *Enhanced and Synthetic Vision 2001*, J. G. Verly, ed., pp. 231–236, Proc. SPIE 4363, 2001.

[6] Z. Rahman, D. J. Jobson, G. A. Woodell, and G. D. Hines, "Multi-sensor fusion and enhancement using the retinex image enhancement algorithm," in *Visual Information Processing XI*, Z. Rahman, R. Schowengerdt, and S. E. Reichenbach, eds., pp. 36–44, Proc. SPIE 4736, 2002.

[7] G. D. Hines, Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multisensor image registration for an enhanced vision system," in *Visual Information Processing XII*, Z. Rahman, R. A. Schowengerdt, and S. E. Reichenbach, eds., pp. 231–241, Proc. SPIE 5108, 2003.

[8] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Spatial aspect of color and scientific implications of retinex image processing," in *Visual Information Processing X*, S. K. Park, Z. Rahman, and R. A. Schowengerdt, eds., pp. 117–128, Proc. SPIE 4388, 2001.

[9] D. J. Jobson, Z. Rahman, and G. A. Woodell, "The statistics of visual representation," in *Visual Information Processing XI*, Z. Rahman, R. A. Schowengerdt, and S. E. Reichenbach, eds., pp. 25–35, Proc. SPIE 4736, 2002. Invited paper.

[10] G. D. Hines, Z. Rahman, D. J. Jobson, and G. A. Woodell, "DSP implementation of the multiscale retinex image enhancement algorithm," in *Visual Information Processing XIII*, Z. Rahman, R. A. Schowengerdt, and S. E. Reichenbach, eds., pp. 13–24, Proc. SPIE 5438, 2004.

[11] G. D. Hines, Z. Rahman, D. J. Jobson, and G. A. Woodell, "Single-scale retinex using digital signal processors," in *Proceedings of the GSPx*, 2004.

Information Theoretic Analysis of Noise Sources in Image Formation

Z. Rahman and D. J. Jobson,

SPIE International Symposium on AeroSense

Visual Information Processing XII

Proceedings SPIE 5108, pp. 39–50

Orlando, FL (2003).

Information theoretic analysis of noise sources in image formation

Zia-ur Rahman[†], Daniel J. Jobson[‡]

[†]College of William & Mary, Department of Applied Science, Williamsburg, VA 23187.
[‡]NASA Langley Research Center, Hampton, Virginia 23681.

ABSTRACT

When a linear image acquisition device captures an image, several noise sources impact the quality of the final (digital) image. We have previously examined the impact of these noise sources in terms of their impact on the total amount of information that is contained in the image, and in terms of their impact on the restorability of the image data. In this paper, we will examine the effect of each of the noise sources on final image quality.

1. INTRODUCTION

The digital image that is produced by the end-to-end imaging process which starts with the capture of light, emitted or reflected by a scene, by a photosensitive device, and which ends with the reproduction of this scene in digital form, contains several artifacts due to the nature of the imaging process. In several previous publications, we have examined the end-to-end imaging system in terms of restorability of the image data acquired by the imaging device, and also in terms of the total information that such a system is capable of transmitting.[1-3] In this paper, we approach the problem from a slightly different perspective: how do these *individual* noise sources affect the *overall* information that the system transmits, and how does this impact the quality of the restored image? We provide graphical results to show the impact of noise sources on the total information and images to show the impact of these noise sources on the final image quality. We, thus, correlate theoretical predictions with actual results using a simulation environment.

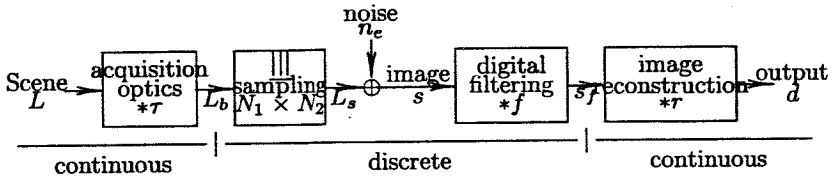


Figure 1: End-to-end imaging system, from (continuous) image scene, to (discrete) signal, and (continuous) display.

For this paper, we are interested only in those artifacts that are part of the image-formation process. We will not examine filters that minimize the impact of these artifacts on the final image. For a discussion of such filters, see, for instance, Huck et al,[1,2] Fales et al,[3] and Park and Rahman.[4] Figure 1 shows the end-to-end imaging process.

1. The light from the (continuous) scene, L, first passes through an optical lens characterized by a point spread function τ to form the (continuous, blurred) representation $L_b = L * \tau$.

2. The blurred image L_b is acquired by the photosensitive CCD array, producing the spatially discrete, $N_1 \times N_2$ sampled image $L_b' = L_b \underline{|||}$, where $\underline{|||}$ is the image sampling lattice given by

$$\underline{|||}(x_1, x_2) = S_1 S_2 \sum_{n_1} \sum_{n_2} \delta(x_1 - n_1 S_1, x_2 - n_2 S_2).$$

   S_1 and S_2 are the intersample distances in the x_1 and x_2 directions respectively. Note that $L_b'[n_1, n_2] = L_b[x_1, x_2]$, $\forall x_1 = n_1 S_1, x_2 = n_2 S_2$. The sampling lattice, in conjunction with the imaging optics, inherently causes aliasing, injecting artifacts into the end-to-end imaging process.[4,5]

3. The (discrete) sampled image, L_b' has continuous amplitude at the sampled nodes. This is converted to digital values with an analog-to-digital (A/D) converter, or quantizer, $\mathcal{Q}$, that produces the digital signal, $L_s = \mathcal{Q}[L_b'] + n_q$, where n_q represents the quantization error that is generated when the continuous sample value is digitized. The quantization operator is given by

$$\mathcal{Q}[y] = \left\lfloor \frac{y 2^{\eta}}{a_{max} - a_{min}} \right\rfloor$$

where η is the number of bits used to represent each sample value and $(a_{max} - a_{min})$ represents the dynamic range of the original scene. If we assume that the dynamic range of the signal has been scaled to match the dynamic range that can be represented by η bits, then $\mathcal{Q}[y] = \lfloor y + 0.5 \rfloor = y + n_q$.

4. In addition to the aliasing, and quantization artifacts, the CCD array also injects artifacts due to thermal and electronic noise into the image. These effects are modeled as additive, Gaussian white noise, n_e. The noise corrupted signal is $s = L_s + n_e$.

5. The discrete, quantized, noisy signal s can be filtered by some operator f to form the discrete, quantized, filtered signal, $s_f = s * f$. Though this operation does not contribute directly to the input artifacts, it may enhance the effects of the artifacts already in the signal. So it does affect the overall quality of the displayed image.

6. The filtered signal s_f is reconstructed by the image display device—monitor, print, or something else—to form the displayed image $d = s_f * r$, where r is the reconstruction filter. The reconstruction filter is typically a low pass filter that blurs the (filtered) signal s_f. For this reason, the filter f typically tends to be a sharpening filter do that the combined response of the reconstruction function and the filter function is (approximately) the ideal low-pass filter.

2. END-TO-END ANALYSIS

There are a total of five sources of artifacts, then, that affect the quality of the final image: optical blurring, aliasing, quantization, thermoelectric noise, and reconstruction blurring. We make the simplifying assumption that these sources are essentially independent of each other so that we can monitor the effect of any one noise source by keeping the rest of them constant. It is difficult to perform this analysis using real images because they already contain these artifacts. Hence, we will use a synthetic image with controllable mean-spatial detail and known power spectral density (PSD) to perform this analysis. Taking a look at the end-end-process *in toto*, the displayed image d can be written as:

$$d(x_1, x_2) = \left\{ \left((L(x_1, x_2) * \tau(x_1, x_2)) \underline{|||}(x_1, x_2) + n_q[n_1, n_2] + n_e[n_1, n_2] \right) * f[n_1, n_2] \right\} * r(x_1, x_2) \quad (1)$$

$$\hat{D}(\omega_1, \omega_2) = \left\{ \left((\hat{L}(\omega_1, \omega_2)\hat{\tau}(\omega_1, \omega_2)) * \hat{\underline{|||}}(\omega_1, \omega_2) + \hat{N}_q[\nu_1, \nu_2] + \hat{N}_e[\nu_1, \nu_2] \right) \hat{F}[\nu_1, \nu_2] \right\} \hat{R}(\omega_1, \omega_2) \quad (2)$$

Notationally, the variables in ( ) are continuous variables, and those in [ ] are discrete. Equation 2 is the Fourier domain representation of Equation 1, and the Fourier transformed variables have a ˆ on top. The (ω_1, ω_2) represent continuous frequencies and, thus, continuous Fourier transform, and the $[\nu_1, \nu_2]$ represent discrete frequencies, and hence discrete Fourier transforms. The Fourier transform of the sampling lattice $\underline{|||}$ is given by

$$\hat{\underline{|||}}(\omega_1, \omega_2) = \sum_{n_1} \sum_{n_2} \delta(\omega_1 - n_1/S_1, \omega_2 - n_2/S_2).$$

The two representations given in Equations 1 and 2 are equivalent, however, the mathematics is much more tractable in the Fourier domain.

Equation 2 can be rewritten to explicitly show all the noise terms that inject artifacts into the displayed image d:

$$\hat{D}(\omega_1,\omega_2) = \left\{\left(\hat{L}_b(\omega_1,\omega_2) * \hat{\underline{\amalg}}(\omega_1,\omega_2) + \hat{N}_q[\nu_1,\nu_2] + \hat{N}_e[\nu_1,\nu_2]\right)\hat{F}[\nu_1,\nu_2]\right\}\hat{R}(\omega_1,\omega_2) \quad (3)$$

$$= \left\{\left(\hat{L}'_b(\omega_1,\omega_2) + \hat{N}_a(\omega_1,\omega_2) + \hat{N}_q[\nu_1,\nu_2] + \hat{N}_e[\nu_1,\nu_2]\right)\hat{F}[\nu_1,\nu_2]\right\}\hat{R}(\omega_1,\omega_2), \quad (4)$$

where $\hat{L}'_b$ is the blurred component of the scene that is encompassed by the sampling passband as determined by the Nyquist frequency, and $\hat{N}_a = \hat{L}_b * \hat{\underline{\amalg}}_o$,

$$\hat{\underline{\amalg}}_o(\omega_1,\omega_2) = \sum_{n_1}\sum_{n_2}\delta(\omega_1 - n_1/S_1, \omega_2 - n_2/S_2), \quad n_1 = n_2 \neq 0.$$

Four of the five sources of artifacts are inherent to the image acquisition process and one, reconstruction blur, is due to the characteristics of image display. Three of the four sources inherent to the image acquisition are modeled as independent, additive noise sources. We will examine the impact of changing these noise terms on the quality of the final image in the next section.

2.1. Metric

One of the measures commonly used to measure the quality of an image is the signal-to-noise ratio (SNR); another is the mutual information, $\mathcal{H}$, between the discrete signal s and the continuous scene L. The mutual information is a function of the SNR so in that sense the two metrics essentially present have the same meaning. However, unlike the generally used definitions of SNR, this particular definition of SNR includes the effects of aliasing in the noise.

The signal s is given as:

$$\hat{s}[\nu_1,\nu_2] = \hat{L}'_b(\omega_1,\omega_2) + \hat{N}_a(\omega_1,\omega_2) + \hat{N}_q[\nu_1,\nu_2] + \hat{N}_e[\nu_1,\nu_2] \quad (5)$$

Assuming stationarity, the PSD of s, $\hat{\Phi}_s = |\hat{s}|^2 = \hat{s}\hat{s}^*$, where $\cdot^*$ represents complex conjugation. The assumption that the noise sources are independent of each other and of the scene, ensures that all of the cross-product terms in the computation of $\hat{\Phi}_s$ are 0. Therefore,

$$\hat{\Phi}_s = \hat{\Phi}_{L_b} + \hat{\Phi}_{n_a} + \hat{\Phi}_{n_q} + \hat{\Phi}_{n_e} \quad (6)$$

$$= \hat{\Phi}_L|\tau|^2 + \hat{\Phi}_{n_a} + \hat{\Phi}_{n_q} + \hat{\Phi}_{n_e} \quad (7)$$

$$= \hat{\Phi}_L|\tau|^2 + \hat{\Phi}_N \quad (8)$$

The right hand side of Equation 8 contains a signal term $\hat{\Phi}_L|\tau|^2$ plus a noise term $\hat{\Phi}_N = \hat{\Phi}_{n_a} + \hat{\Phi}_{n_q} + \hat{\Phi}_{n_e}$. One can then compute the mutual information between the acquired signal and the scene. From Huck et al,[6–8] we see that the mutual information $\mathcal{H}$ is determined by the ratio $\hat{\Phi}_s/\hat{\Phi}_N$. Thus,

$$\mathcal{H} = \frac{1}{2}\iint_{\hat{B}} \log_2\left[1 + \frac{\hat{\Phi}_s(\omega_1,\omega_2)}{\hat{\Phi}_N(\omega_1,\omega_2)}\right] d\omega_1\, d\omega_2 \quad (9)$$

$$= \frac{1}{2}\iint_{\hat{B}} \log_2\left[1 + \frac{\hat{\Phi}_L(\omega_1,\omega_2)|\hat{\tau}(\omega_1,\omega_2)|^2}{\hat{\Phi}_{N_a}(\omega_1,\omega_2) + \hat{\Phi}_{N_q}[\nu_1,\nu_2] + \hat{\Phi}_{N_e}[\nu_1,\nu_2]}\right] d\omega_1\, d\omega_2 \quad (10)$$

$$= \frac{1}{2}\iint_{\hat{B}} \log_2\left[1 + \frac{\hat{\Phi}_L(\omega_1,\omega_2)|\hat{\tau}(\omega_1,\omega_2)|^2}{\hat{\Phi}_L(\omega_1,\omega_2)|\,\hat{\tau}(\omega_1,\omega_2)|^2 * \hat{\underline{\amalg}}_o + \hat{\Phi}_{N_q}[\nu_1,\nu_2] + \hat{\Phi}_{N_e}[\nu_1,\nu_2]}\right] d\omega_1\, d\omega_2. \quad (11)$$

The mutual information can be computed as a function of the blurring, aliasing, quantization, and thermoelectric noise artifacts, as long as we know the PSD $\hat{\Phi}_L$ of the scene. Following Itakura et.al[9] and Kass and Hughes,,[10] we assume that the PSD of the radiance field L is

$$\hat{\Phi}_L(\omega_1,\omega_2) = \frac{2\pi\mu^2\sigma_L^2}{\left[1+\left(2\pi\mu(\omega_1^2+\omega_2^2)\right)^2\right]^{3/2}}, \tag{12}$$

which is circularly symmetric. In Equation 12, μ is the mean spatial detail in the (stationary) scene, and σ_L is the standard deviation. This power spectral density has shown to closely match the power spectral density of natural scenes.[9] Additionally, we assume that the thermoelectric noise and the quantization are white Gaussian, so that:

$$\hat{\Phi}_{N_e} = \sigma_e^2 \tag{13}$$

$$\hat{\Phi}_{N_q} = \sigma_q^2 \approx \left(2^{-b}\sigma_s\right)^2, \tag{14}$$

$$\sigma_s^2 = \iint \hat{\Phi}_s(\omega_1,\omega_2)\,d\omega_1 d\omega_2 \approx \iint \hat{\Phi}_L(\omega_1,\omega_2)|\hat{\tau}(\omega_1,\omega_2)|^2\,d\omega_1 d\omega_2. \tag{15}$$

The optical transfer function (OTF) of the combined lens and CCD optics can be (closely) modeled as a Gaussian.[1,11] For this reason, we assume that

$$\hat{\tau}(\omega_1,\omega_2) = \exp\left[-(\omega_1^2+\omega^2)/\sigma_\tau^2\right],$$

where σ_τ is the parameter that controls the width of the Gaussian, and hence directly determines the trade-off between aliasing and blurring (see Section 3.2). Figure 2 shows the OTF for various values of σ_τ. The sampling passband is also shown for comparison.

3. INFORMATION AND SOURCES OF ARTIFACTS

Using the PSD defined in Equation 12 in conjunction with the definition of mutual information, $\mathcal{H}$, we can examine he impact of the noise sources on the total information that the system acquires. In Section 4, we will correlate these noise artifacts with image quality.

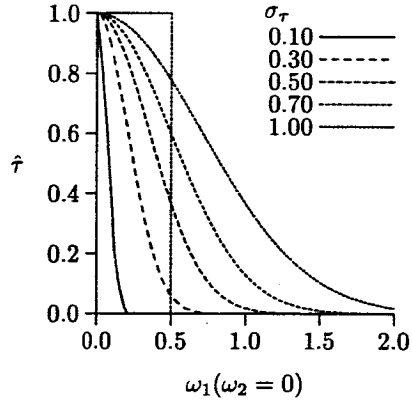


Figure 2. The OTF $\hat{\tau}$ of the imaging optics for various values of the control parameter σ_τ. The sampling passband for unit sampling interval, $S_1 = S_2 = 1$, is shown in dashed lines. As σ_τ increases, the OTF blurs less, but more and more of the (blurred) signal falls outside the sampling passband, giving rise to greater aliasing.

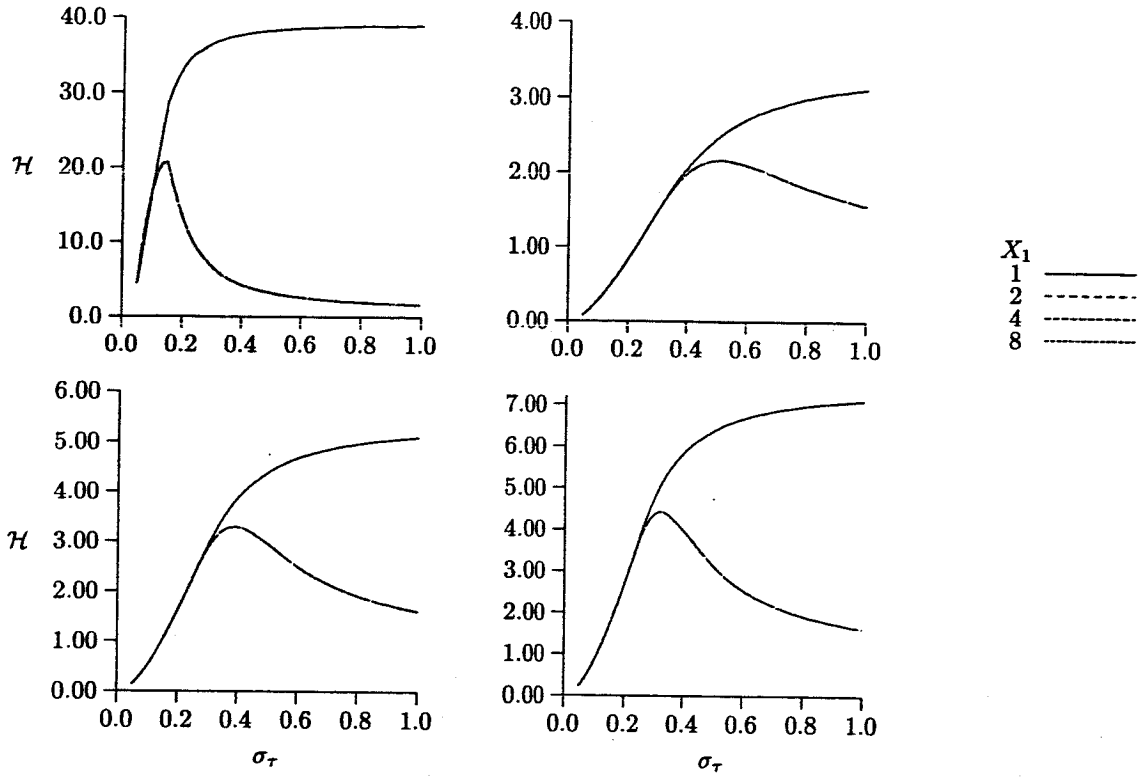


Figure 3. The mutual information $\mathcal{H}$ as a function of the OTF $\hat{\tau}$ of the imaging optics and aliasing. The information is maximized when the imaging system can balance the increased noise due to aliasing against the loss of information due to blurring. Top left: SNR = ∞; top right: SNR = 16; bottom left: SNR = 64; bottom right: SNR = 256.

3.1. Optical blurring

The optical blurring due to the combined response of the lens and the image-acquisition lattice is modeled by the OTF $\hat{\tau}$. This is not an actual noise source, thought it is source of artifacts especially when σ_τ is very small. In fact, it can be shown that in the absence of any noise, the effects of blurring can be completely removed by applying an "inverse" filter to the blurred image. Mathematically, since $\hat{s} = \hat{L}\hat{\tau}$ in the absence of noise, then if $\hat{F} = (\hat{\tau})^{-1}$, then $\hat{s_f} = \hat{s}\hat{F} = (\hat{L}\hat{\tau})(\hat{\tau})^{-1} = \hat{L}$. This would be true for all values of σ_τ if the images are stored with infinite precision. However, due to the A/D conversion, the inverse filter cannot undo the blurring for those values where the signal has been quantized to 0. It was determined (experimentally) that for an 8-bit A/D converter, signals blurred with $\sigma_\tau \geq 0.125$ can be completely recovered; those blurred with $\sigma_\tau < 0.125$ can be either recovered partially or not at all (Figure 7 bottom row). However, imaging systems are rarely, if ever, noise free. In such cases, applying the inverse filter to the image can serve to actually enhance the noise artifacts.[12] So while blurring does not contribute directly to the noise, its overall impact is to make the image less contrasty and sharp. Figure 2 shows how the change in the σ_τ parameter affects the OTF.

3.2. Aliasing

Where the effects of optical blurring contribute directly to noise is when sampling is introduced into the analysis. The traditional trade-off in the design of the acquisition part of the imaging system is really a trade-off between aliasing and blurring in the presence of noise. Since we are assuming for this analysis that the thermoelectric and quantization noise terms are constant, we can also assume, without loss of generality, that they are identically 0. Hence the only noise contribution is due to aliasing which, in turn, depends completely on the OTF of the combined responses of the optics and image-gathering lattice. Figure 3 shows the impact of varying the

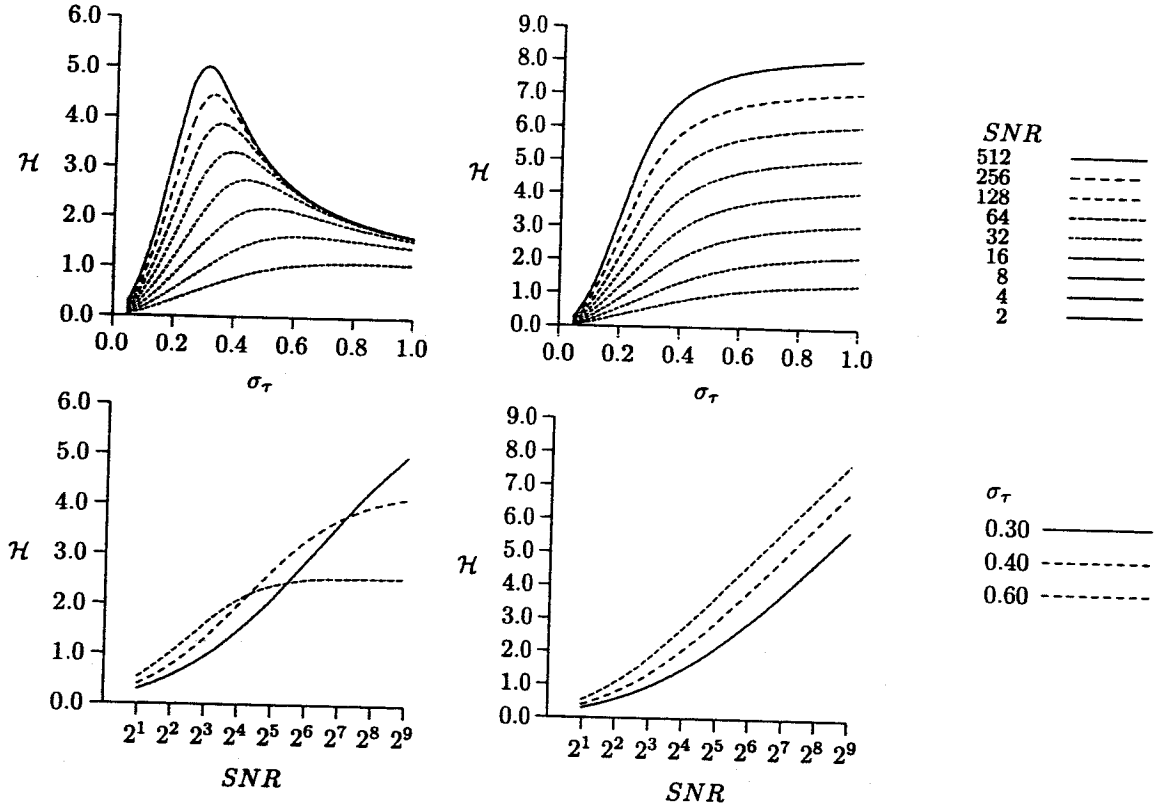


Figure 4. The mutual information $\mathcal{H}$ as a function of the thermoelectric signal-to-noise. Right column: $mu = 1; X_1 = 4$; left column: $\mu = 1; X_1 = 1$.

OTF and intersample distance X_1 on the total information. The figure shows that as the OTF gets wider, the amount of aliasing increases causing the total information to decrease. Conversely, when the blurring increases, aliasing decreases, leading to increased information. However, past a certain threshold, blurring further reduces the total amount of information. The solid line represents the case when there is no aliasing or noise and the only limiting factor is the image blur. Additional results are shown for the cases where there is aliasing and thermoelectric noise in the system.

3.3. Thermoelectric noise

The thermoelectric noise is modeled as white, Gaussian. Since the PSD of a white Gaussian function is a constant, any increase in the noise term leads directly to a decrease in the total information $\mathcal{H}$. Figure 4 shows the impact of changing the SNR on the total information. For the left column, the intersample distance is set to 4 so there is aliasing as well as thermoelectric noise. The right column shows the same thing except in the absence of aliasing. The top-left graph can be used to derive the σ_τ parameter that maximizes information for a given SNR. So given the imager's thermoelectric SNR, the lens can be designed so that the optics and the sampling array combine to maximize information. As SNR decreases, more and more aliasing can be tolerated since it does not affect the overall quality of the image more than the thermoelectric noise. The top-right graph shows that if aliasing is not the limiting factor, then the information continues to increase as more and more signal is passed through the lens unblurred. The bottom row shows the trend as a function of the SNR for given aliasing parameters σ_τ. Information increases monotonically with increasing SNR. However, at low SNRs, the signals with higher aliasing contain more information because they are sharper. At high SNR, the signals with less aliasing contain more information because the aliasing becomes the major noise source.

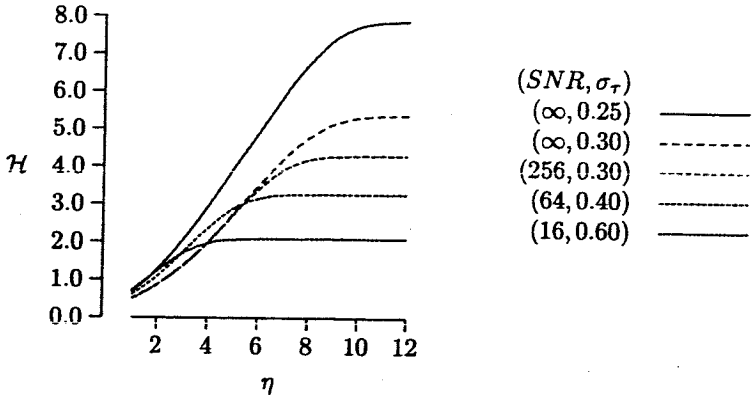


Figure 5. The mutual information $\mathcal{H}$ as a function of the quantization rate η. The information increases as the quantization rate increases.

3.4. Quantization

The quantization process is an interesting one. Since the idea is to represent the dynamic range of the (continuous) phenomenon being observed as a bounded (discrete) set, there are a couple of trade-offs that need to be understood. If we want to capture the full dynamic range of the phenomenon, then unless the range is bounded, i.e. beyond a given upper and lower limit, the values are essentially zero, it is difficult to establish the upper and lower limits that are needed to define the quantizer. If an incorrect model is used, then the quantization error can be significant. If however we do assume a bounded source, as is the case here since we are measuring the amount of light that is passing through the lens and then falling on the CCD array, we can assume different types of quantization models and use the one that provides the least error. We can show that if we assume that the light obeys a Gaussian distribution, then the quantization error obeys a uniform distribution with the error given in Equation 15.[1] We have previously examined the impact of quantization on sampling-limited systems.[13] Figure 5 shows the impact of quantization on total information assuming all other sources remain constant. Results are shown for five cases: the first two represent the situation when there is no thermoelectric noise: the image quality is then restricted by the blurring function τ and the quantization rate. The last three show the cases for those designs that maximize information by matching the thermoelectric noise with the appropriate blurring function parameter σ_τ. The information increases with the quantization rate. However, beyond a particular quantization rate η_l, increasing η does not yield additional information. The value of η_l depends upon the combination of the amount of thermoelectric noise and the blurring.

4. IMAGE QUALITY AND SOURCES OF ARTIFACTS

In Section 3 we described the (theoretical) metric of mutual information $\mathcal{H}$ with the hypothesis that maximizing the metric gives the best image quality. In this section we are going to examine the actual effects of the artifacts on the overall (visual) image quality. We do this by varying the different parameters that are used to characterize the various sources of artifacts. the best to way to examine the the effects of the artifacts on image quality, is to use a synthetic image with known characteristics, such as PSD and mean spatial detail. We use a computer generated scene of random polygons following the procedure by Modestino and Fries.[14] This target has definable mean spatial detail, and its PSD closely matches that given in Equation 12. Since this is a computer generated image it is free from image formation artifacts. Note that we say "image formation" artifacts only; the generated images is not free from *all* imaging artifacts because of any artifacts that are particular to the display device. For instance, reconstruction and interpolation artifacts are always going to be present when we display the synthetic image, and it always has to be displayed: otherwise it is just a matrix of integers. Figure 6 shows the synthetic target for mean spatial detail $\mu = 1$, but with varying zoom factors.

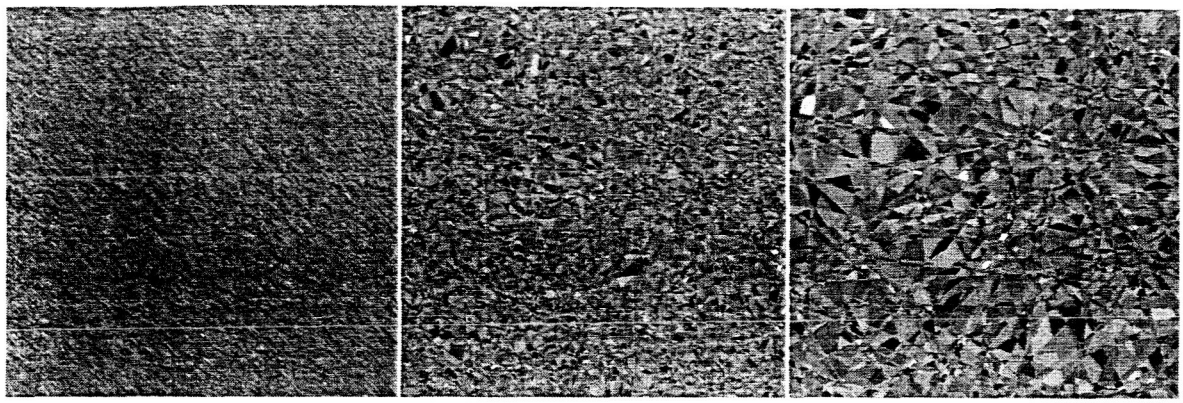


Figure 6. Synthetic target with $\mu = 1$. The left image is at full resolution, the middle at a zoom factor of 4, and the right at a zoom factor of 8. The zoom factors represent the effective sampling density

4.1. Blurring

The blurring artifacts are easiest to visualize, and display. In order to show only blurring, we assume that the image is sufficiently sampled, i.e., the PSD of the image lies completely within the Nyquist sampling passband. We can do this in the synthetic simulation environment but this assumption is not valid for any real imaging system. One of the impacts of sufficiently sampling a signal is that the overall appearance of the reconstructed image is blurry: the image could be made sharper by allowing some aliasing (see Figure 8. The original scene can be completely recovered from this blurry representation by using an inverse filter, as long as $\hat{\tau}(\omega_1, \omega_2) > 0, \quad \forall(\omega_1^2 + \omega_2^2)^{1/2} < \hat{B}$, where $\hat{B}$ is the Nyquist sampling frequency. Figure 7 shows the impact of blurring on the original image and also the restored image when the appropriate inverse filter is applied.

4.2. Aliasing

Because we are examining the impact of artifacts in a simulation environment, we can examine the impact of aliasing with and without a lens in the imaging system. The effects of aliasing are most easily visualized in the frequency domain. Aliasing is a direct consequence of sampling. The process of sampling in the spatial domain is equivalent to the process of replication in the frequency domain. As long there is sufficient distance between the replicas in the frequency domain, the image can be fully reconstructed from its frequency domain representation. Since there is an inverse relationship between the intersample distance in the spatial and frequency domains, this implies that the image has to be sampled on a fine lattice in the spatial domain. When this is not the case, there is crossover energy between the neighboring replicas. This effect is known as aliasing. Since the lens applies a Gaussian blur to the scene, it, in effect, attenuates the signal in a given profile determined by the shape of the OTF. The "blurrier" the OTF, the less aliasing there is in the final image. This is because blurry lenses attenuate the signal rapidly, causing most of the signal that passes through to fall within the Nyquist passband. Conversely, the less blurry the OTF, the greater the aliasing. Figure 8 shows the impact of aliasing with and without a lens in place. If the lens is designed so that there is no overall aliasing—the OTF is designed so that the PSD of the signal passing through the lens is confined (almost) completely to the Nyquist passband—then the original signal can be recovered with the application of an inverse filter. In such a case, there is no aliasing, only blurring. However, most real systems build in aliasing, because the overall image is then sharper. If the systems were designed with digital post-filtering, then they could be designed with considerable blurring with the OTF because the impact can be reversed with digital filtering. However, because of thermoelectric noise (Section 3.3) this filtering can be less than desirable. Therefore, the systems are designed with aliasing to produce as sharp an image as possible without obvious artifacts. This relates back to the trade-off between aliasing and blurring that was mentioned in Section 3.2.

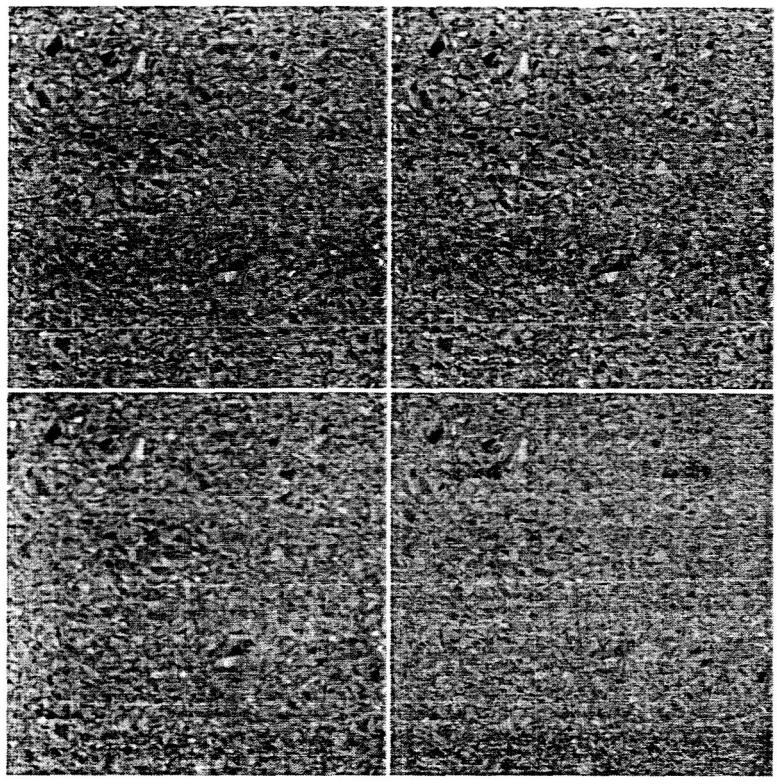


Figure 7. The impact of blurring without aliasing in the absence of noise on the overall image quality. The original image can be completely recovered from the blurred image by applying the appropriate inverse filter as long as $\sigma_\tau > 0.125$—top row. For scenes acquired with lenses with $\sigma_\tau < 0.125$, information cannot be completely recovered—bottom row..

4.3. Thermoelectric noise

Though the thermoelectric noise source is modeled as additive, white Gaussian, it poses one of the biggest challenges for image quality. In the *absence* of thermoelectric noise, an inverse filter can be used to recover from blurring and, to some extent, aliasing artifacts. However, when there is additive thermoelectric noise present in the image, the rules for restoring the image change considerably. A typical inverse filter has low magnitude at high frequencies. The noise in the frequency domain has a constant magnitude at all frequencies. Hence, the inverse filter magnifies the impact of noise at high frequencies. Filters have to be designed carefully in order to minimize the impact of thermoelectric noise on the restored image..[1,2] Figure 9 shows the impact of thermoelectric noise on the overall image quality in the absence of other noise sources. The PSD $\hat{\Phi}_{n_e}$ due to n_e has to be quite high before its impact is clearly visible in the output image, especially in the case of printed media where the signal is dithered—i.e., has random noise added to it—before reconstruction.

4.4. Quantization

As mentioned earlier, quantization impacts are usually very small unless very coarse quantization is applied. In figure 10 we show the synthetic scene for several quantization levels. As the quantization grows coarser, the amount of information drops directly as a function of the quantization rate, (Section 3.4). This holds true for the visual quality as well, though the visual system is quite tolerant to fairly large quantization error. For most *printed* images, a quantization rate of $\eta = 5$ gives satisfactory performance depending upon the type of print media—plain or photographic quality paper, or transparency—, printing process—inkjet, thermal dye sublimation, etc.—, and the printing medium—ink, ribbon, etc. For most CRT displays, a quantization rate

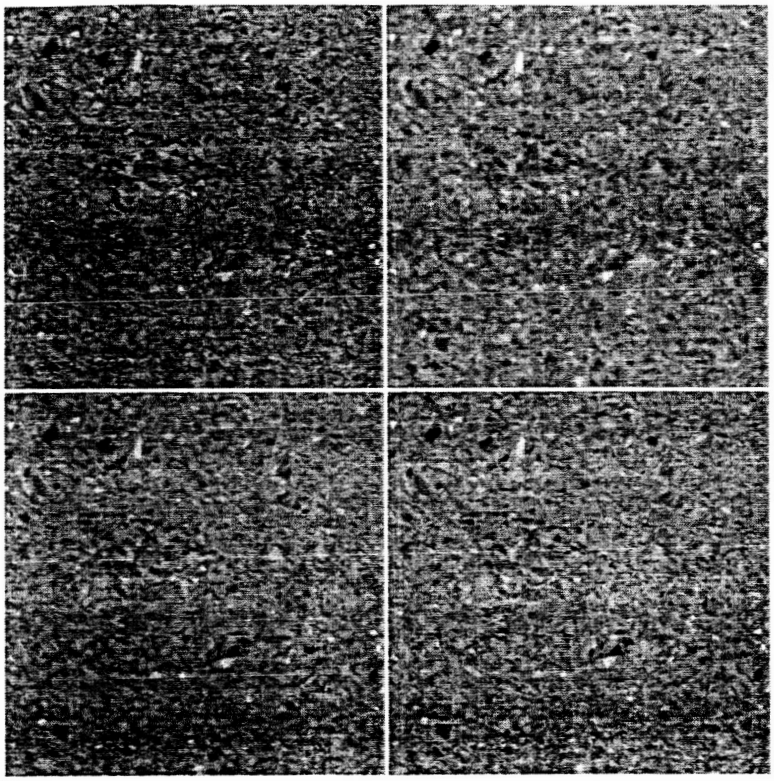


Figure 8. The impact of aliasing—with and without a (blurring) lens. When the lens attenuates the signal so that it is (almost) zero outside the Nyquist passband, the acquired signal can be recovered with an inverse filter. Otherwise, aliasing causes the image to appear sharper, but also introduces artifacts that are most noticeable in scenes with high low mean spatial detail, i.e. scenes with large areas of constant intensity. These effects are most noticeable as edge replication—or ringing–near edges or large changes in intensity.

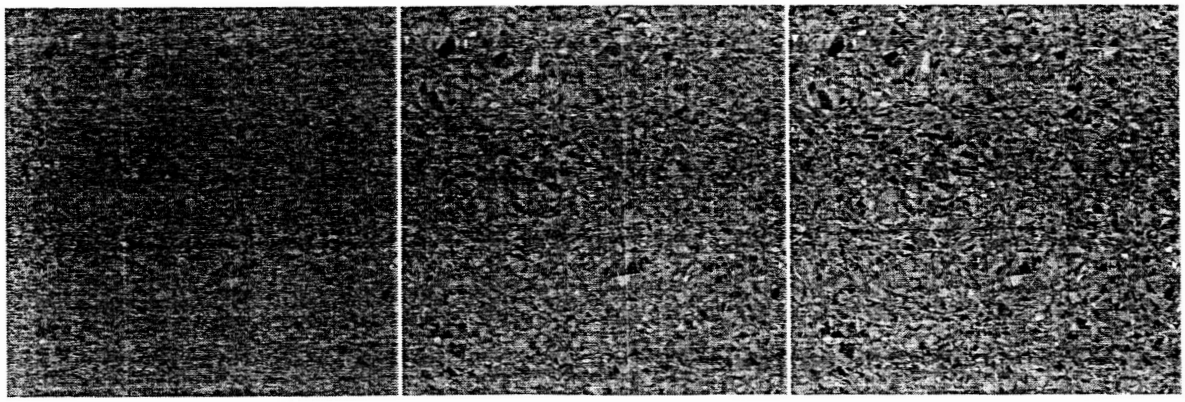


Figure 9. The impact of thermoelectric noise: high noise (left), medium noise (middle), low noise (right). The PSD $\hat{\Phi}_{n_e}$ due to n_e has to be considerable before the impact on visual quality is visible.

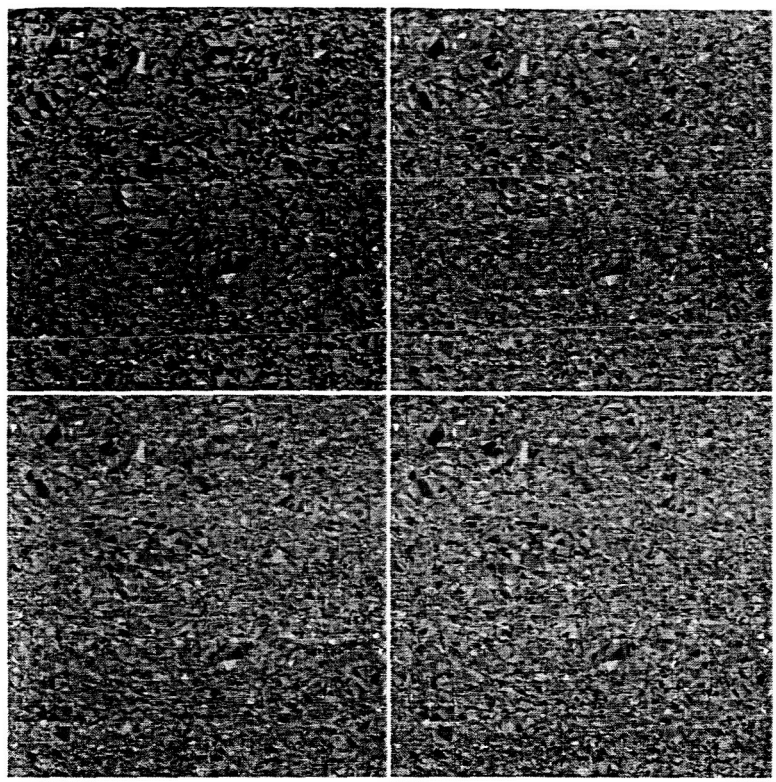


Figure 10. The impact of quantization on image quality. Results are shown for $\eta = 8$: bottom right; $\eta = 6$: bottom left; $\eta = 4$: top right; and $\eta = 2$: top left. As the quantization gets coarser, the artifacts such as false contouring become more obvious. However, there is little discernible difference between $\eta = 4, 6$, and 8.

of $\eta = 6$ or $\eta = 7$ is sufficient. Since most digital display device quantize data at $\eta = 8$, quantization artifacts rarely have an impact on the overall quality of the image, but should be considered for completeness.

5. CONCLUSIONS

We have examined image acquisition artifacts with respect to their impact on the amount of mutual information that the device can transmit and the overall image quality. We examined four noise sources: optical blurring,aliasing, thermoelectric noise and quantization. Though we examined the sources individually, we showed that they do have an impact on each other. For instance, the effects of aliasing and blurring are intimately intertwined, as are the effects of thermoelectric noise and blurring. We showed that there is a strong correlation between the hypothesis that best results are obtained when information is maximized and the overall image quality.

We have not really addressed the question of how these various observations can be combined to design the very best imaging system. However, we have shown the performance bounds due to the various noise sources, and also, to some extent, the interrelationships between the noise sources. These can be used as a recipe to design a system that can be used to maximize both the information and image quality.

Acknowledgments

Dr. Rahman's work was supported with the NASA cooperative agreement NCC-1-01030, under the aegis of the Synthetic Sensors Element of the Aviation Safety Program.

Contact Information

Zia-ur Rahman zrahman@as.wm.edu
Daniel J. Jobson daniel.j.jobson@nasa.gov

REFERENCES

1. F. O. Huck, C. L. Fales, and Z. Rahman, *Visual Communication: An Information Theory Approach*, Kluwer Academic Publishers, Norwell, MA, 1997.
2. F. O. Huck, C. L. Fales, and Z. Rahman, "Information theory of visual communication," *Philosophical Transactions of the Royal Society of London A* **354**, pp. 2193–2248, Oct. 1996.
3. C. L. Fales, F. O. Huck, R. Alter-Gartenberg, and Z. Rahman, "Image gathering and digital restoration," *Philosophical Transactions of the Royal Society of London A* **354**, pp. 2249–2287, Oct. 1996.
4. S. K. Park and Z. Rahman, "Fidelity analysis of sampled imaging systems," *Optical Engineering* **38**, pp. 786–807, May 1999.
5. S. K. Park and R. Hazra, "Aliasing as noise: A quantitative and qualitative assessment," in *Visual Information Processing II*, pp. 2–13, Proc. SPIE 1961, 1993.
6. F. O. Huck and S. K. Park, "Optical-mechanical line-scan imaging process: Its information capacity and efficiency," *Applied Optics* **14**(10), pp. 2508–2520, 1975.
7. F. O. Huck, N. Halyo, and S. K. Park, "Information efficiency of line-scan imaging mechanisms," *Applied Optics* **20**(11), pp. 1990–2007, 1981.
8. F. O. Huck, C. L. Fales, N. Halyo, R. W. Samms, and K. Stacy, "Image gathering and processing: Information and fidelity," *Journal of the Optical Society of America A* **2**(10), pp. 1644–1666, 1985.
9. Y. Itakura, T. Suteo, and T. Takagi, "Statistical properties of the background noise for the atmospheric windows in the intermediate infrared region," in *Infrared Physics*, **14**, pp. 17–29, Pergamon Press, New York, NY, 1974.
10. M. Kass and J. Hughes, "A stochastic model for AI," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 369–372, IEEE, (New York, NY), 1983.
11. F. O. Huck, C. L. Fales, D. J. Jobson, and Z. Rahman, "Electro-optical design for efficient visual communication," *Optical Engineering* , March 1995.
12. S. K. Park and R. Hazra, "Image restoration versus aliased noise enhancement," in *Visual Information Processing III*, pp. 52–62, Proc. SPIE 2239, 1994.
13. Z. Rahman, "Information capacity of sampling-limited systems," in *Integrated Computational Imaging Systems*, Optical Society of America, 2001.
14. J. W. Modestino and R. W. Fries, "Construction and properties of a useful two-dimensional random field," *IEEE Transactions on Information Theory* **26**, pp. 44–50, 1980.

Multisensor Image Registration For An Enhanced Vision System

G. D. Hines, Z. Rahman, D. J. Jobson and G. A. Woodell

SPIE International Symposium on AeroSense

Visual Information Processing XII

Proceedings SPIE 5108, pp. 231-241

Orlando, FL (2003)

Multi-image registration for an enhanced vision system

Glenn Hines[a], Zia-ur Rahman[b], Daniel Jobson[a],Glenn Woodell[a]

[a]NASA Langley Research Center, Hampton, VA 23681;
[b]College of William & Mary, Department of Applied Science, Williamsburg, VA 23187

ABSTRACT

An Enhanced Vision System (EVS) utilizing multi-sensor image fusion is currently under development at the NASA Langley Research Center. The EVS will provide enhanced images of the flight environment to assist pilots in poor visibility conditions. Multi-spectral images obtained from a short wave infrared (SWIR), a long wave infrared (LWIR), and a color visible band CCD camera, are enhanced and fused using the Retinex algorithm. The images from the different sensors do not have a uniform data structure: the three sensors not only operate at different wavelengths, but they also have different spatial resolutions, optical fields of view (FOV), and bore-sighting inaccuracies. Thus, in order to perform image fusion, the images must first be co-registered. Image registration is the task of aligning images taken at different times, from different sensors, or from different viewpoints, so that all corresponding points in the images match. In this paper, we present two methods for registering multiple multi-spectral images. The first method performs registration using sensor specifications to match the FOVs and resolutions directly through image resampling. In the second method, registration is obtained through geometric correction based on a spatial transformation defined by user selected control points and regression analysis.

Keywords: image registration, control points, resampling

1. INTRODUCTION

Improving aviation safety is a primary mission of the National Aeronautics and Space Administration (NASA). Synthetic Vision Systems (SVS) is an initiative at the NASA Langley Research Center (LaRC) to improve the level of aviation safety. One component of the SVS is the Enhanced Vision Systems (EVS) project where several sensor technologies are being utilized to provide vision-enabling information to commercial and general aviation pilots to assist them in poor visibility conditions.

Infrared sensors are routinely used in remote sensing applications. Coupling an infrared sensor with a visible band sensor — for frame of reference or for additional spectral information — and properly processing the two information streams has the potential to provide valuable information in night and/or poor visibility conditions. A multi-spectral infrared and visible EVS[1] was developed at LaRC to test this concept. This system contains a short wave infrared (SWIR) camera, a long wave infrared (LWIR) camera, a color, visible band, charge coupled device (CCD) camera, and the hardware and software to process the data streams from the cameras. The system has been deployed and flown on the NASA LaRC B757 ARIES research aircraft.

The digitized images obtained from the data streams of the visible and infrared cameras are processed using the Multi-scale Retinex (MSR) algorithm. The MSR is a general-purpose image enhancement algorithm for producing good visual representations of scenes. It performs a non-linear spatial/spectral transform that synthesizes strong local contrast enhancement and color constancy.[2,3] Though it was developed for color CCD images, the MSR has been applied to infrared data from the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) and the LANDSAT thematic mapper to pre-process the images for better classification.[4,5] The dynamic range compression, color constancy and sharpening provided by the MSR makes it an ideal algorithm to use for image enhancement in poor weather conditions. Since the MSR is used to enhance infrared as well as visible spectral bands, different features in different bands can be accentuated. From this a composite or fused image can be created that contains more information than any individual spectral band. For the EVS, the MSR serves as both an enhancement processor and fusion engine.[6,7] A set of images consisting of an image from each of the cameras taken during one time-aligned frame is fused into a single image. This process is then repeated for all the image frames making up a video sequence.

To properly perform fusion it is critical to ensure that the information from each sensor refers to the same features in the environment.[8,9] The different sensors of the EVS have different acquisition lattices and optics, therefore they capture information in data structures that are substantially different from each other. Thus, the images must first be registered before any fusion is performed. Several authors[10–12] have addressed image registration problems with innovative, but often complex, general solutions. In this paper we describe two straightforward solutions for registering the EVS images.

The first solution is a registration algorithm based on simply correcting camera sensor specification differences through cropping, scaling and image resampling, assuming that the cameras are bore-sighted and hence share a common center. We will show and discuss an example of registration with this algorithm. Tests based upon this method of registration showed that the assumption that the cameras were perfectly bore-sighted was invalid and adjustments had to be made to correct for the differences in sensor orientation and alignment. A second algorithm was implemented that factored in these obvious distortions and gave better results than the original algorithm. However, whereas the first algorithm was completely automatic — the only inputs were the characteristics of the different sensors and which sensor provided the baseline — the second algorithm does require a manual selection of control points. It should be pointed out, though, that if this correction was to be performed on a video sequence, then only one frame is needed for manual selection of control points. The coefficients that are obtained from this correction can be used to correct all the other frames, assuming that the nature of the distortion does not change. We will also show and discuss examples of this algorithm.

2. BACKGROUND

2.1. Image registration

Image registration is the task of aligning images taken at different times, from different sensors, or from different viewpoints so that all corresponding points in the images match. A transform must be defined that relates the points in one image to their corresponding points in another. This transform depends upon the characteristics of the differences between the images being registered, and is computed with respect to a reference or baseline image. The images that are to be matched to the reference are called the sensed, or, distorted image.

More particularly, image registration is defined as a mapping between two or more images both spatially (geometrically) and with respect to intensity. Expressed mathematically we have:

$$I_2 = g\big(I_1(f(x_1, x_2))\big),$$

where I_1 and I_2 are two-dimensional images (indexed by x_1, x_2), $f : (x_1, x_2) \rightarrow (x_1', x_2')$ maps the indices of the distorted frame to match those of the reference frame, and g is a one-dimensional intensity or radiometric transform.[13] In this paper, we assume that we do not need to make any radiometric adjustments, so $g = \mathcal{I}$, the identity transform. Hence we are concerned only with the spatial transformation, f. In generating a spatial transform for the EVS, our primary difficulty is the lack of fiducial markers within the images generated by the EVS sensors. The cameras are, however, assumed to be bore-sighted so they are expected to have a common center of alignment. The spatial transform should, then, properly align the images, but should not affect any characteristic differences that should be exposed by registration.

Spatial transforms may take on different forms depending upon the application. Simple, common transforms specified by analytic expressions include rigid-body, affine, projective or perspective, and polynomial.[14,15] The distortions between the images of the EVS in general seem constrained to those correctable by affine transforms. They also appear to be characterizable by a global (versus local[16]) transform where a single transform correctly maps all the points on the distorted image to match the corresponding points on the reference image. An affine transform fulfills the requirements for the needed transform.

An affine transform can perform rotation, translation, scaling and shearing operations. It offers six degrees of freedom when selecting six unknown coefficients and solving a system of six linear equations. In general, it can

perform triangle-to-triangle mappings. A general representation of an affine transform is $[y_1, y_2, 1] = [x_1, x_2, 1]T$ where

$$T = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix},$$

x_1 and x_2 reference the input coordinate system, y_1 and y_2 reference the output coordinate system, and a_{ij} are transform coefficients.[17]

The forward mapping functions are

$$y_1 = a_{11}x_1 + a_{21}x_2 + a_{31}$$

and

$$y_2 = a_{12}x_1 + a_{22}x_2 + a_{32}.$$

Geometric, image-to-image registration can be summarized in three general steps:

1. Feature identification and matching is performed to establish a correspondence between features in the distorted image to those in the reference image;
2. A spatial transformation is selected and the transformation coefficients are computed based upon the feature matching criteria;
3. The distorted image is inverse-mapped using the computed transformation and resampled to register it with the reference image.

Feature identification and matching are often performed by selecting pixel locations called control points. Identification of control points can be accomplished in several different ways.[16,18] Manual identification of control points is commonly performed. The images are displayed, normally side-by-side, and corresponding points usually based on features such as lines, edges, or contours are selected from both images.

The spatial transform coefficients that represent the unknown image distortions are determined from the control points. A minimum of three non-collinear control points are required to determine the six unknown coefficients of an affine transformation. Wolberg and Jensen[17,19] describe several techniques to solve for unknown coefficients including pseudo-inverse solutions, least squares with ordinary and orthogonal polynomials, and weighted least squares with orthogonal polynomials.

Image resampling is the process of transforming a sampled image from one (input pixel grid) coordinate system to another (output pixel grid), where a sampled image is the digitization of the spatial coordinates of an image function $f(y_1, y_2)$ — a two-dimensional intensity function.[17,20,21] The two coordinate systems are related to each other by the mapping function of a spatial transformation.

To perform image resampling, initially, the output pixels are inverse mapped using the transformation function to a new grid which (usually) doesn't correspond to the input grid. Thus an interpolation (image reconstruction) procedure is used to generate a continuous surface through the samples of the new grid. Then the input image is sampled (digitized) at these points to provide the discrete output pixel values of the process. Three common methods of interpolation are nearest neighbor, bilinear, and parametric cubic convolution.[17,22]

2.2. Sensor specifications

The cameras chosen for the EVS are commercial-off-the-shelf (COTS) devices. The SWIR camera sensor captures wavelengths between 1-3 μm. The LWIR (thermal) camera sensor captures wavelengths between 8-12 μm. And, the color visible band CCD camera operates between 0.4-0.7 μm. Table 1 shows the relevant manufacturer characteristics of the sensors.[1] The images from the three sensors obviously need to be be registered because of the differences in these characteristics. The solutions developed to resolve these differences are discussed in Section 3.

	SWIR	LWIR	CCD
Image Dimensions (pixels)	$320H \times 240V$	$320H \times 240V$	$542H \times 497V$
Optics FOV	$34°H \times 25°V$	$39°H \times 29°V$	$34°H \times 25°V$
Detector Readout Frame Rate	60Hz (typical)	60Hz	30Hz (interlaced)

Table 1: Sensor Specifications

The characteristics of the actual images obtained for registration differ from the initial manufacturer specifications because of data acquisition and storage to tape. First, all images have a nominal image size of 640 x 480 pixels corresponding to the NTSC format of the recorded images. However, the actual size of the images is quite different after the images are cropped so that the FOVs match the "visible" part of the images (see Section 3). Second, ground test measurements of the cameras' FOVs differed from the manufacturer provided values. These updated characteristics are shown in Table 2, and need to be included in the computations for proper registration of the data streams.

	SWIR	LWIR	CCD
Image Dimensions (pixels)	$640H \times 480V$	$640H \times 480V$	$640H \times 480V$
Optics FOV	$31.5°H \times 23.5°V$	$41°H \times 30.75°V$	$33.5°H \times 25°V$
Detector Readout Frame Rate	60Hz (typical)	60Hz	30Hz (interlaced)

Table 2: Updated Sensor Specifications

2.3. Algorithm overview

The algorithms operate on a set of three, time-aligned images where each image is acquired by an individual camera of the EVS. Each of the video streams is recorded, or post-processed, with video timecode information in each frame. The frames are time-aligned simply by finding the frames with matching time codes. This set of time-aligned frames is then used to obtain the registration parameters with respect to the baseline frame. All other frames of the video sequence can be processed with the same parameters. Each frame, including the ones from the color CCD sensor, is converted to grayscale before registration and further processing.

3. REGISTRATION ALGORITHMS

Our first solution for image registration algorithm is based solely on camera sensor specifications. The cameras were assumed to be properly bore-sighted at installation thus the only distortion parameters to account for in registration are the differences in FOVs and resolutions. This algorithm, called the SS (sensor specifications) algorithm, performs registration by first equalizing the FOVs and then resampling the distorted image to match reference resolutions. Based upon the lessons learned from the SS algorithm, a geometric image-to-image registration algorithm was implemented. Both of these algorithms are discussed below. For each of the algorithms, we use the SWIR image as the baseline since it has the "worst" image parameters (the smallest FOV and poorest spatial resolution). The size of an image can be modified through interpolation but we cannot increase the FOV.

3.1. SS algorithm

The first step of the SS algorithm is to equalize the instantaneous FOVs (IFOV)s of the sensors. The FOV is the angular extent of the full image on the sensor and the IFOV is the angular extent on an individual detector element, i.e., the solid angle through which a detector element is sensitive to radiation.

From Figures 1, 2, and 3, we observe that the visible portion of the images is actually smaller than the full image capture window. The FOVs listed in Table 2 are assumed to correspond to the visible portion and

not the capture window. Thus, the first stage of processing is to crop the images to the visible portions. The second stage of processing is to ensure that the two images are representing the same portion of the scene. Since the FOVs of the SWIR and the LWIR sensors differ — LWIR has the greater FOV and hence captures a wider swath of the scene — the LWIR image needs to be cropped so that it encompasses the same FOV as that encompassed in the SWIR image. The dimensions of the cropped LWIR images — the number of columns and rows — are determined by a simple scaling operation. The horizontal and vertical IFOVs of the LWIR image are obtained using

$$\texttt{IFOV-LWIR-HORIZONTAL} = \frac{\texttt{FOV-LWIR-HORIZONTAL}}{\texttt{LWIR-COLS}}$$

and

$$\texttt{IFOV-LWIR-VERTICAL} = \frac{\texttt{FOV-LWIR-VERTICAL}}{\texttt{LWIR-ROWS}}$$

respectively. The number of cropped columns and rows for the LWIR image is then determined by

$$\textbf{columns} = \frac{\texttt{FOV-SWIR-HORIZONTAL}}{\texttt{IFOV-LWIR-HORIZONTAL}}$$

$$\textbf{rows} = \frac{\texttt{FOV-SWIR-VERTICAL}}{\texttt{IFOV-LWIR-VERTICAL}}$$

After cropping, the SWIR and LWIR FOVs are equal, but since the dimensions of the cropped LWIR are different from the dimensions of the SWIR, the IFOVs of the LWIR and the SWIR images are still different. To make the IFOVs the same we must resample the cropped LWIR image so that it is the same size as the SWIR image. This entails: (1) computing an expansion factor that will make the dimensions of the cropped LWIR image greater than the dimensions of the SWIR (2) pixel replicating the cropped LWIR based on the expansion factor and (3) downsampling the expanded LWIR image to the SWIR dimensions. We use the bi-linear interpolation method.[23] Nearest neighbor interpolation can also be selected if desired but bilinear interpolation is more spatially accurate and results in images that are slightly smoother.

A similar sequence of operations is performed between the SWIR image and the visible image. If the FOVs are the same, as in Table 1 then the visible image is simply downsampled to match the SWIR resolution. The initial results from the SS algorithm clearly indicated that the distortions present in the images were not excessive, but they also were not limited to FOV and resolution differences.

3.2. MLR algorithm

Based on the results obtained from the SS algorithm a more general, geometric image to image registration algorithm is implemented. The distortions between the images seem to be due to sensor translation, (slight) rotation, scale change, and, possibly, shear. An affine transform is, thus, used to model the spatial transformation. Control points are manually selected for identifying and matching corresponding features between the reference and distorted images. Since we assume that the sensors do not change alignment over time, we only need to register one baseline set of images that can subsequently be used for the rest of the image frames.

We use point mapping without feedback[13] to approximate the global affine transformation. The first stage of the MLR algorithm is to select a minimum of three non-collinear control points from two input images. More points can be chosen to make the coefficients more representative of the distortions throughout the overall image if the points are well distributed. Global distortion representation is also improved by choosing pixels on the perimeter if possible. The control points are then analyzed using multiple linear regression[24,25] to approximate the coefficients of the affine transform. Residuals to determine the accuracy of the regression model obtained are calculated. The defined affine transform provides a mapping between the baseline and distorted images. The distorted image is then resampled using the transform parameters to create the registered image. Bilinear interpolation is used for resampling.

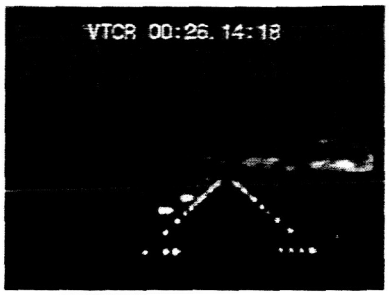


Figure 1: Original SWIR

Figure 2: Original LWIR

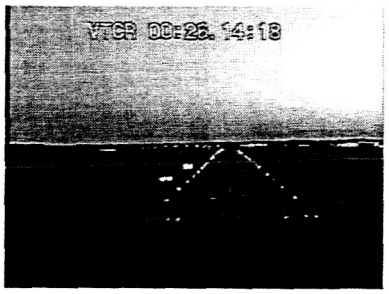


Figure 3: Original Visible

4. RESULTS

To demonstrate the performance of the algorithms we processed a set of videos taken by the EVS cameras during a flight test at Patrick Henry airport in Newport News, Virginia. The video sequence was of the B757 aircraft approach to a runway, and was digitized using a Canopus Video Board. Three images (one from each camera) time-aligned at 00:26:14:18 were used for registration. As stated earlier, the SWIR image is used as the baseline for registration since it has the poorest spatial resolution and FOV. The SWIR, LWIR and visible images are shown in Figures 1, 2 and 3 respectively.

To provide a similarity metric to validate the performance of the registration algorithms we take the absolute difference of the reference and corrected images. This provides a visible validation of the registration process since features such as runway edges should align if registration is performed correctly.

4.1. SS algorithm

Applying the SS algorithm with the SWIR image as the baseline, and the LWIR and visible images as distorted images yields the "registered" SWIR, LWIR and visible images shown in Figures 4, 5 and 6 respectively. The FOV of the LWIR image has been made smaller to match the FOV of the SWIR image. This change in FOV can clearly be seen in the horizontal direction of Figure 5, by observing that the blurred artifact (which is an antenna in the FOV of the camera) in the upper left corner of the original LWIR is now almost completely removed in the registered image. In the vertical direction, the decrease in FOV is noted by the missing timecode at the top and the missing ground features at the bottom of Figure 5 that are in the original image. The IFOVs have also been matched though resampling. The general effects of resampling can be seen by observing the expansion of image features from Figure 2 to Figure 5. The FOV of the original visible image in Figure 3 has been made slightly smaller, again to match the FOV of the SWIR, in Figure 6. Since the FOVs nearly match and the image dimensions are the same, there is only a small expansion to match IFOVs, hence the registered image features are only slightly increased from the original.

Figure 7 is the differenced SWIR and SS registered LWIR, and Figure 8 is the differenced SS registered LWIR and SS registered visible image. The misalignment between the images after registration can clearly be seen in Figure 7 by observing the difference in the outline of the runway from the LWIR component of the image, and the runway lights from the SWIR image. There is at least a large translation and a small rotation difference between the SWIR and registered LWIR. Similarly, the misalignment between the registered LWIR and visible images differenced in Figure 8 can also be seen by noting the difference in the outline of the runway from the LWIR image, and the runway lights from the visible image. Again, there is an obvious translation between the images. Figures 7 and 8 clearly display the misalignment between the images thus indicating that differences in sensor design characteristics are not the only cause of distortion between the images.

4.2. MLR algorithm

We first applied the MLR algorithm to the original (uncropped) SWIR and visible images, again using the SWIR as the baseline. Due to the lack of features around the perimeter of the SWIR image we used the runway lights as control points. Note that we are only using three control points for demonstration purposes. Figure 9

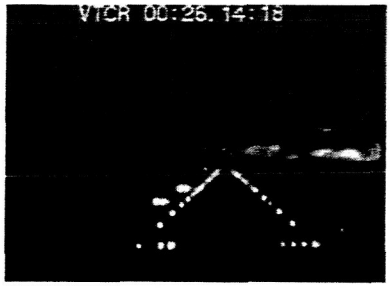


Figure 4: Cropped SWIR

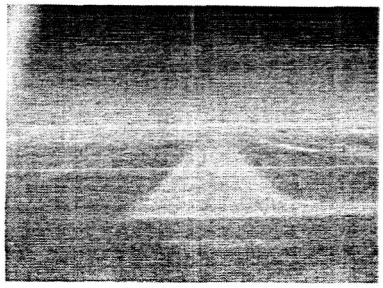


Figure 5: SS Reg. LWIR

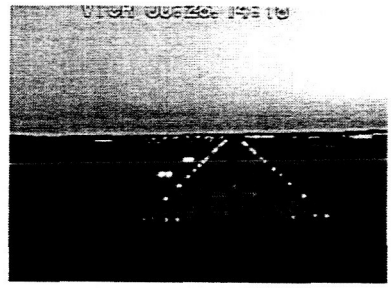


Figure 6: SS Reg. visible

Figure 7: SWIR and SS Registered LWIR

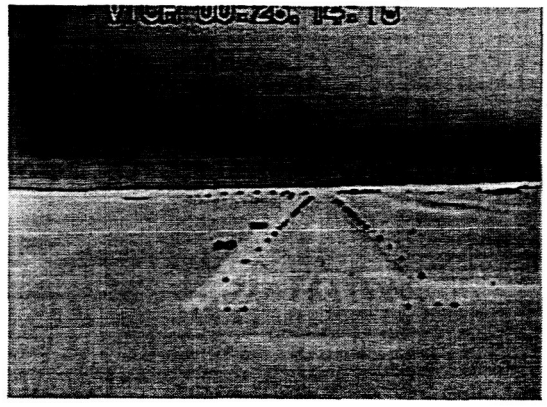


Figure 8: SS Registered LWIR and visible

repeats the original SWIR image for reference. Figure 10 shows the registered visible image and Figure 11 is the differenced SWIR and registered visible image. The coefficients obtained are given in Table 3.

	b_0	b_1	b_2
x'	-0.546156	1.021212	-0.004578
y'	-20.440557	-0.007477	0.972837

Table 3: Visible to SWIR MLR Coefficients

A close look at the runway and the runway lights in the two images shows that they are now registered. In particular, in the lower right corner of the SWIR image there are four runway lights lined up horizontally. In the visible image there are three runway lights in the same position, except the second light from the left is not visible. Figure 11 shows the four lights differenced in a horizontal line with the missing visible light filled in from the SWIR image. It is clear to see the warp performed during registration by observing the timecode size and location differences in the differenced images. The timecodes are the same size and at the same location in the original images. Figure 9 and Figure 10 could now be equally cropped to remove disjoint pixels around the perimeter to obtain the final images to be fused.

Next we applied the MLR algorithm to the registered visible and LWIR images using the visible image as the baseline. Since the runway lights are not visible in the LWIR image, we use the intersecting lines at the bottom and top of the runway, and a stripe at the beginning of the runway towards the right in the LWIR image as control points. Figure 12 repeats the MLR registered visible image for reference. Figure 13 is the registered LWIR image and Figure 14 is the differenced registered visible and the just registered LWIR images. The coefficients obtained are given in Table 4.

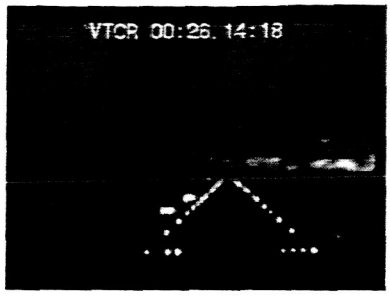


Figure 9: Repeated SWIR

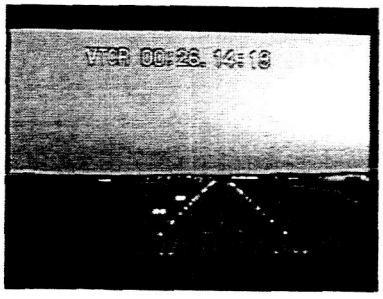


Figure 10: MLR Reg. visible

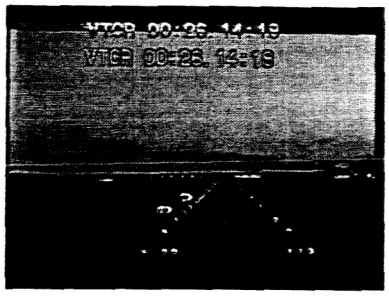


Figure 11. SWIR and MLR Reg. visible

	b_0	b_1	b_2
x'	8.347350	0.850628	0.037684
y'	9.637629	-0.012015	0.779082

Table 4: LWIR to visible SWIR MLR Coefficients

As is evident in Figure 14 the runway portion of the LWIR image is aligned with the runway portion of the registered visible image. Also, the runway lights from the visible image border the perimeter of the LWIR runway. The one runway stripe selected as a control point is aligned. The taxiways on the right side of the image and the horizon across the image are also aligned. Again, any disjoint pixels around the perimeter could be removed by cropping. At this point all three original images are registered.

As a final test of the MLR algorithm we applied the same control point coefficients to a later frame in the video sequence. Figures 15, 16 and 17 are the SWIR, LWIR and visible images at time 00:26:14:28, 10 seconds later in the sequence. Figure 18 shows the MLR registered visible image. Figure 19 is the differenced SWIR and MLR registered visible images. Figure 20 is the MLR registered LWIR image. Figure 21 is the differenced MLR registered visible and MLR registered LWIR image. As in the previous set of images, the registration can be observed by noting the alignment of the runway and runway lights in Figures 19 and 21.

4.3. Discussion

The registration images shown visually demonstrate the performance of the two algorithms on typical image data from the EVS. The registration inaccuracies of the SS algorithm are obvious. The differing FOV and resolution specifications given do not take into account the other distortions within the images. With this much discrepancy there seems to be either a fundamental problem in the bore-sighting or alignment of the cameras, or the alignment is changing during flight. If the sensors were actually bore-sighted and aligned, the

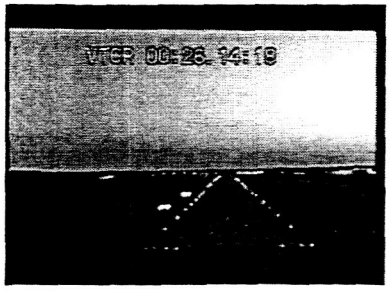


Figure 12. Repeated MLR Registered visible

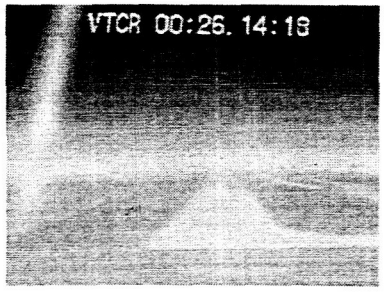


Figure 13: MLR Reg. LWIR

Figure 14. MLR Reg. visible and LWIR

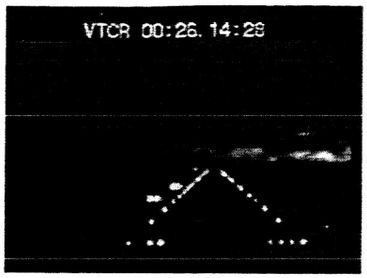


Figure 15. Orig. SWIR at Time 26:14:28

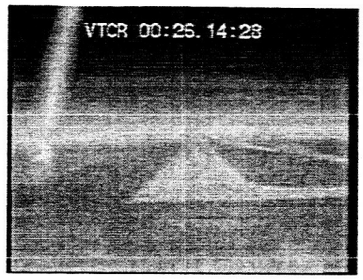


Figure 16. Orig. LWIR at Time 26:14:28

Figure 17. Orig. visible at Time 26:14:28

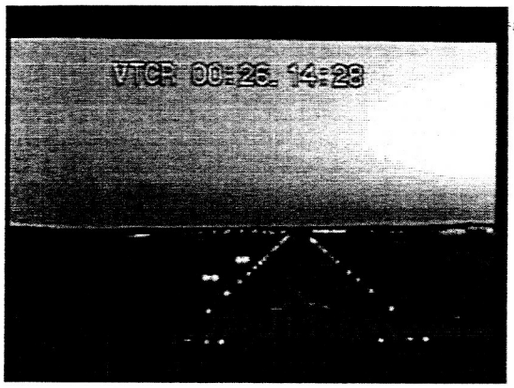


Figure 18. MLR Registered visible at Time 26:14:18

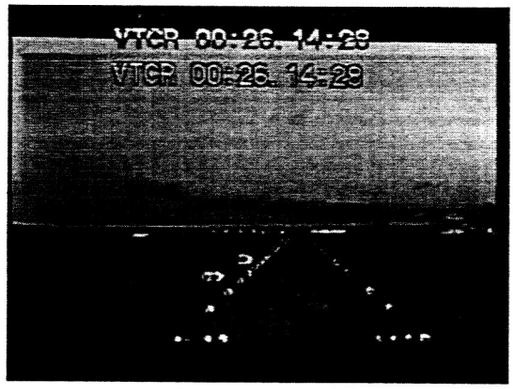


Figure 19: SWIR and MLR Registered visible

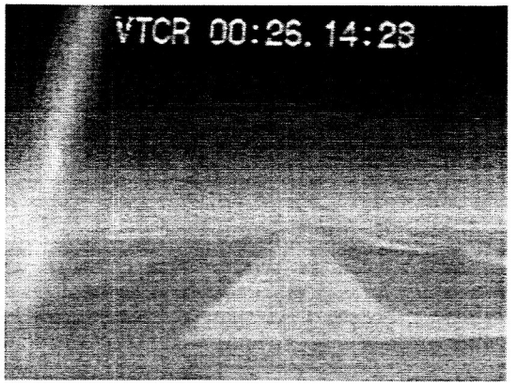


Figure 20. MLR Registered LWIR at Time 26:14:18

Figure 21: MLR Registered visible and LWIR

SS algorithm should be able to match the performance of the MLR algorithm and in addition, not require any manual intervention like selection of control points. True FOV values could be obtained from a thorough ground calibration, and non-interpolated pixels of the actual image dimensions could be obtained from raw digital data streams from the cameras.

The MLR algorithm provides better registration of the images than the SS algorithm configured with the current set of specifications and with the current EVS alignment. In our examples the runway and runway lights are clearly aligned. The coefficients obtained with only three points indicates that there are rotation, translation, scale and possibly shear distortion components found between the images. These distortions can be seen by viewing the timecode warps at the top of the differenced images. The application of the same MLR control points to a set of time-aligned images later in the same video sequence produced the same level of registration. This indicates that we could successfully use the registration coefficients obtained from one set of time-aligned images to apply to, at least, a group of frames from the video sequence. If the alignment is not changing substantially during flight then all frames could be processed with the same transform.

5. CONCLUSIONS

Image registration is an essential prerequisite to subsequent image fusion. This research effort produced two algorithms to perform multi-image registration for the EVS. The SS algorithm uses EVS camera specifications and performs registration based solely on these parameters. The performance of this algorithm indicates that there is a severe inaccuracy in the boresighting or alignment of the cameras. Correction of these issues should improve the performance of the algorithm and allow it to be used to automatically register all images across the cameras, or as validation of the MLR algorithm.

The MLR algorithm uses control point selection and linear regression to compute the coefficients of an affine spatial transformation. This transformation is then used to register the LWIR and visible images to the SWIR image. In addition, the MLR registration algorithm provides a means to generate a base set of coefficients for post processing of the full video stream across all cameras. We have subsequently used a set of baseline coefficients to process an entire 20 second video clip from each of the three cameras.

In addition, the coefficients obtained could also be used to back out the actual distortion values (translation amount, rotation angle, etc.) for feedback to the EVS designers. Improvements could also be made in the computation of the coefficients by using point selection with feedback or other more robust feature selection mechanisms. Manual control point selection can be improved by MSR enhancement of the images to emphasize and sharpen features prior to registration. This was done for another EVS data set and greatly improved the ability to select corresponding points. Most importantly, the actual boresighting and alignment can be checked against the values obtained from MLR and SS registration, and adjusted appropriately. This procedure could be performed both before, and after, EVS flight opportunities and used to verify and validate system alignment.

REFERENCES

1. C. Tiana, J. Kerr, and S. Harrah, "Multispectral uncooled infrared enhanced vision system for flight test," in *Proceedings of SPIE*, **4363**, April 2000.
2. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing* **6**, pp. 451–462, March 1997.
3. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965–976, July 1997.
4. B. D. Thompson, Z. Rahman, and S. K. Park, "Retinex preprocessing for improved multi-spectral image classification," in *Visual Information Processing VIII*, Proc. SPIE 3716, 1999.
5. B. D. Thompson, Z. Rahman, and S. K. Park, "A multi-scale retinex for improved performance in multi-spectral image classification," in *Visual Information Processing IX*, Proc. SPIE 4041, 2000.
6. Z. Rahman, D. Jobson, G. Woodell, and G. Hines, "Multi-sensor fusion and enhancement using the retinex image enhancement algorithm," in *Visual Information Processing XI, Proceedings of SPIE 4736*, April 2002.

7. Rahman. see http://dragon.larc.nasa.gov/fog_haze for examples.
8. R. Brooks and S. Iyengar, *Multi-Sensor Fusion: Fundamentals and Applications*, Prentice Hall, 1998.
9. R. Luo and M. Kay, *Multisensor Integration and Fusion for Intelligent Machines and Systems*, Ablex, Norwood, NJ, 1995.
10. H. Li and Y. Zhou, "Automatic visual/ir image registration," in *Optical Engineering*, **35(2)**, pp. 391–400, February 1996.
11. R. Sharma and M.Pavel, "Multisensor image registration," in *SID Digest, Society for Information Display*, **XXVIII**, pp. 951–954, May 1997.
12. H. Li, B. Manjunath, and S. Mitra, "A contour-based approach to multisensor image registration," in *IEEE Transactions on Image Processing*, **4, No. 3**, March 1995.
13. L. Brown, "A survey of image registration techniques," in *ACM Computing Surveys*, **24, No.4**, December 1992.
14. R. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Academic Press, 1997.
15. K. Novak, "Rectification of digital imagery," in *Photogrammetric Engineering and Remote Sensing*, **58, No. 9**, pp. 399–344, March 1992.
16. A. Goshtasby, "Image registration by local approximation methods," in *Image Vision Computing*, **6**, pp. 255–261, November 1988.
17. G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, 1990.
18. L. Fonseca and B. Manjunath, "Registration techniques for multisensor remotely sensed imagery," in *Photogrammetric Engineering and Remote Sensing*, **62, No. 9**, pp. 1046–1056, September 1996.
19. J. Jensen, *Introductory Digital Image Processing*, Prentice Hall, 1996.
20. N. Dodgson, "Image resampling," Tech. Rep. 261, University of Cambridge, United Kingdom, August 1992.
21. R. Gonzalez and R. Woods, *Digital Image Processing*, Addison-Wesley, 1993.
22. S. Park and R. Schowengerdt, "Image reconstruction by parametric cubic convolution," in *Computer Vision, Graphics, and Image Processing*, **23**, pp. 258–272, 1983.
23. H. Burdick, *Digital Imaging*, McGraw Hill, 1997.
24. G. Snedecor, *Statistical Methods*, Iowa State University, 8 ed., 1989.
25. D. Wittink, *The Application of Regression Analysis*, Allyn and Bacon, 1988.

Feature visibility limits in the non-linear enhancement of turbid images

D. J. Jobson, Z. Rahman, and G. A. Woodell,

SPIE International Symposium on AeroSense

Visual Information Processing XII

Proceedings SPIE 5108, pp. 24–30

Orlando, FL (2003)

Feature visibility limits in the non-linear enhancement of turbid images

Daniel J. Jobson[‡], Zia-ur Rahman[†], Glenn A. Woodell[‡]

[‡]NASA Langley Research Center, Hampton, Virginia 23681.
[†]College of William & Mary, Department of Computer Science, Williamsburg, VA 23187.

ABSTRACT

The advancement of non-linear processing methods for generic automatic clarification of turbid imagery has led us from extensions of entirely passive multiscale Retinex processing to a new framework of active measurement and control of the enhancement process called the Visual Servo. In the process of testing this new non-linear computational scheme, we have identified that feature visibility limits in the post-enhancement image now simplify to a single signal-to-noise figure of merit: a feature is visible if the feature-background signal difference is greater than the RMS noise level. In other words, a signal-to-noise limit of approximately unity constitutes a lower limit on feature visibility.

1. INTRODUCTION

In the previous development of non-linear image enhancement methods,[1–3] our goal was to enhance the visual realism of the recorded digital image to more closely approach the generally much better visibility of direct scene perception by the human observer. For images acquired under turbid—fog, smoke, haze, snow, rain—imaging conditions, there is already a close parity between the recorded image and the direct observation. So the goal of enhancement now becomes fundamentally different: we wish to greatly exceed the performance of the human observer. This is of particular interest for enhancing imagery acquired under turbid aviation conditions. A generic automatic computation that does this provides the enabling technology for real-time image enhancement that can be projected to the pilot's heads-up display (HUD). This type of imagery is especially important in commercial aviation during runway approach and landing and in general aviation during the entire flight sequence from take-off to landing.

Our interest in enhancing images acquired under turbid imaging conditions, coupled with the scientific insights[4] gained from previous purely passive retinex processing led us to formulate a more comprehensive framework of active measurement and control of the image enhancement process: the Visual Servo (VS). The major lesson learned from these scientific insights was that the good visual representations produced by retinex processing all converged uniquely to an ideal statistical characterization. This, together with additional constraints, led to the formulation of an entirely new set of visual measures for image contrast, lightness and sharpness. These measures form the basis for VS controls that affect the level of image enhancement and, hence, image quality. The VS additionally contains a special module for detecting turbid imaging conditions and invokes special processing to produce maximal scene feature clarity.

This framework represents a form of visual intelligence: the software quantitatively assesses visual quality before and after each enhancement step, and guided by these measurements, strives to achieve a standard high level of visual quality. The VS controls still rely on non-linear image processing elements, so conventional end-to-end systems analyses[5–8] cannot be employed to characterize the imaging and computing scheme as a whole. Therefore we seek to understand how to characterize performance in lieu of having linear systems analysis as a tool. In this paper we describe the Visual Servo concept, present results for diverse turbid imaging conditions to indicate its generic performance as an automatic computation, and examine the critical issue of defining a figure of merit for the post-enhancement feature visibility limit in turbid imaging conditions.

Contact: DJJ: d.j.jobson@larc.nasa.gov; ZR: zrahman@as.wm.edu; GAW: g.a.woodell@larc.nasa.gov;

2. THE VISUAL SERVO (VS) CONCEPT

Our extensive previous experience with retinex image enhancement,[3,9] led us to conclude that further improvements in enhancement were possible in terms of contrast and image sharpness and that the extreme narrow dynamic range case of turbid imaging could be also be encompassed by a fundamental shift in approach away from purely passive retinex processing to an active measurement and control system. The scientific insights[4] gained from the experience of retinex processing of very large numbers of highly diverse images provided the foundations for constructing entirely new visual measures for image contrast, lightness, and sharpness. Underlying this effort is the core idea that there is an ideal visual representation with consistent statistics for recorded images[3,4] and that the enhancement process is one of trying to make any image approach this ideal as closely as possible. This core idea also relates to the visual inadequacy of the linear representation of recorded image data. The idea rejects the notion of imaging as a replication process with quality defined by minimizing artifacts, and shifts to the notion of imaging being a (highly non-linear) transformation process that seeks to achieve a good visual representation whose statistics depart sharply from those of the linear representation.

With this background, the study of the overall global statistics of good visual representations revealed that global statistics alone could not support the definition of visual measures.[4] There was insufficient capture of the visual sense of contrast and lightness. However, regional spatial ensembles did provide a basis for quantifying visual contrast and lightness. These were augmented by the development of a new sharpness measure which together capture the most comprehensive and key visual elements of images. The measures have been extensively tested, but are not yet ready for full technical exposition, so here we will discuss them and the resulting VS at the conceptual and schematic level.

The Visual Servo is shown in Figure 1. The reason for calling the computation a "servo" stems from the fact that it is based upon similar ideas to electro-mechanical servo systems of active measurement and control. The basic flow for enhancing an image is as follows:

1. measure a key visual parameter
2. based upon the measured value, activate an enhancement control to improve the overall brightness, contrast and sharpness of the image
3. recompute the visual measure
4. if the measured value of the parameter achieves the high visual standard then terminate the process
5. otherwise, iterate until either the visual standard has been achieved or the VS has determined internally that all reasonable enhancement processing has been exhausted

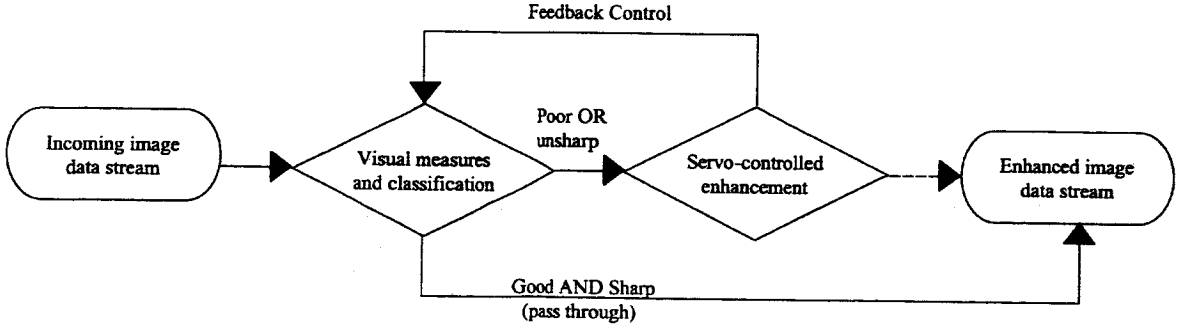


Figure 1: The automatic visual servo.

Images acquired under turbid imaging conditions represent a special case for detection and processing due to their extremely narrow, but unpredictable, dynamic range. For these images, the VS determines when this

very low contrast is occurring, and then invokes custom processing to achieve a very powerful enhancement. This enhancement, from the results to be shown in Figure 2, appears to produce maximum feature contrast that is limited only by the noise inherent to the imaging process.

3. VISUAL SERVO RESULTS FOR TURBID IMAGING

Turbid imaging covers a very wide range of imaging conditions where there is obscuration between the scene and the imaging sensor due to particle scattering in the imaging medium. For atmospheric turbidity, the veiling can be due to fog, smoke, haze, dust, rain, or snow, or some combination of these, such as smog. For underwater imaging the turbidity is most often due to suspended cellular plant life, sediment particles, or some combination thereof. In order to validate the generality of the computation for turbid imaging conditions, we have tested the VS on images with very varied types and degrees of turbidity, as well as highly diverse scene content. The result is an automatic, general purpose computation that is as applicable to aerial imagery as it is to underwater imagery. All of the results shown in Figure 2 (right column) were obtained in the fully automatic mode, without any additional tweaking of parameters or post-processing. Figure 2 (middle column) shows the performance of the default MSRCR on the same images. It is evident from comparing the two columns why we moved away from the passive MSRCR to the active VS for enhancing images acquired in turbid imaging conditions.

The turbidity detection and custom processing of the VS has worked extremely well in (almost) all the cases that we have tested thus far. For lightly turbid conditions, the VS transitions smoothly to the non-turbid enhancement processing, invoking different servo modules that provide "weaker" enhancements. The VS has been tested with many hundreds of still color images, as well as with color and long, and short wave infrared (IR) video imagery. The IR imagery was acquired from aviation sensors during test flights, or from sensors mounted on a 250 feet high gantry to simulate flight conditions such as long throw views of landing approaches.

The performance of the VS has been outstanding for still and video imagery.* Additionally, it does not seem to be sensitive to the type of particulate scattering involved. Good clarity—visibility distance improvement—was achieved for moderate fog, severe haze, moderately thick smokes, heavy rain and snow as well as for moderately thick underwater turbidities. Some clarity was possible for heavy fog conditions (see Figure 4). This latter performance limitation in dense fog was true for all imagery types tested—color, short wave IR, and long wave IR. For all but the dense fog case, sufficient clarity was achieved so that often all traces of obscuration were removed. For cases of thicker turbidity, visibility as an increase in feature distance visibility was greatly improved. For the cases where we acquired the image data ourselves, we could compare the performance of the VS to what we had observed at the time of acquisition. In all cases, except severe fog, the VS result was far better than our observed visibility.

A reasonable baseline for performance comparison is to compare servo results with those for the conventional automatic histogram modification method—autolevels. Autolevels is a moderately powerful automatic image enhancement technique which is quite useful for images with low contrast that do not contain regions of saturation. It is a histogram stretch technique that adjusts the dynamic range of the displayed image based upon a fixed parameter that determines the significant dynamic range of the input image. The dynamic range compression adjustment of the Autolevels process is very different from the intrinsic dynamic range compression that the non-linear processing embedded in the VS provides. So the primary performance differences to be expected are the result of the non-linear dynamic range compression. Figure 3 shows examples of a comparison between the performance of Autolevels and the VS. It is clear from the figure that the VS has much better performance than Autolevels. This implies that even for very narrow dynamic range imaging, the non-linear dynamic range compression is still quite advantageous. The reason for this is that the very narrow dynamic range is not stable regionally across the image. There is still lot of "shading" variation due to spatial lighting effects within the narrow dynamic range of the turbid image. Of course there can be cooperative cases where the lighting happens to be spatially uniform, and for these cases, the Autolevels performance will approach that of the VS. The servo performance though should be much more all encompassing of the full complexity of real turbid imaging conditions where shadows, and other lighting variations, as well as highly spatially variable degrees of turbidity are going to be encountered within a particular image frame.

*Further examples of VS enhancements can be found on `http://dragon.larc.nasa.gov/retinex`.

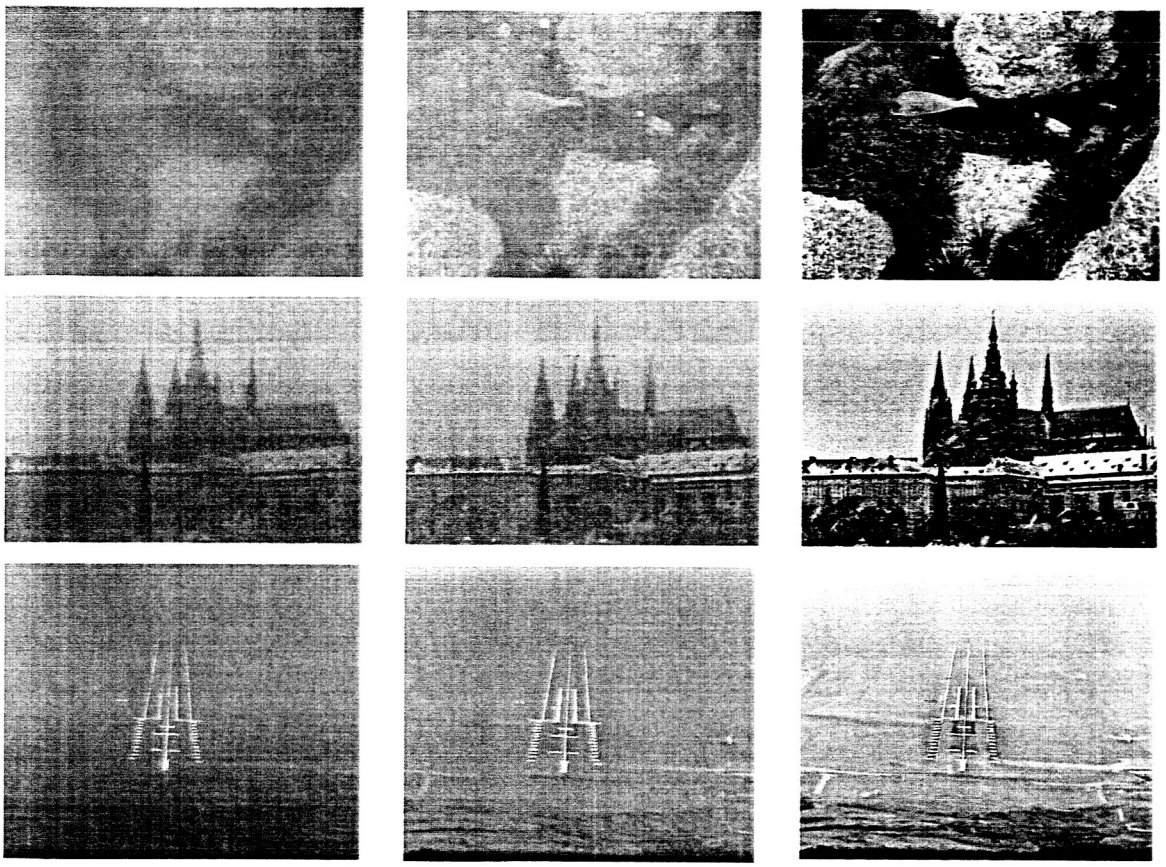


Figure 2. The performance of the VS and the default MSRCR on images acquired under turbid imaging conditions. The left column shows the original images; the middle column the default MSRCR processed images; and the right column, images that have been processed using the VS.

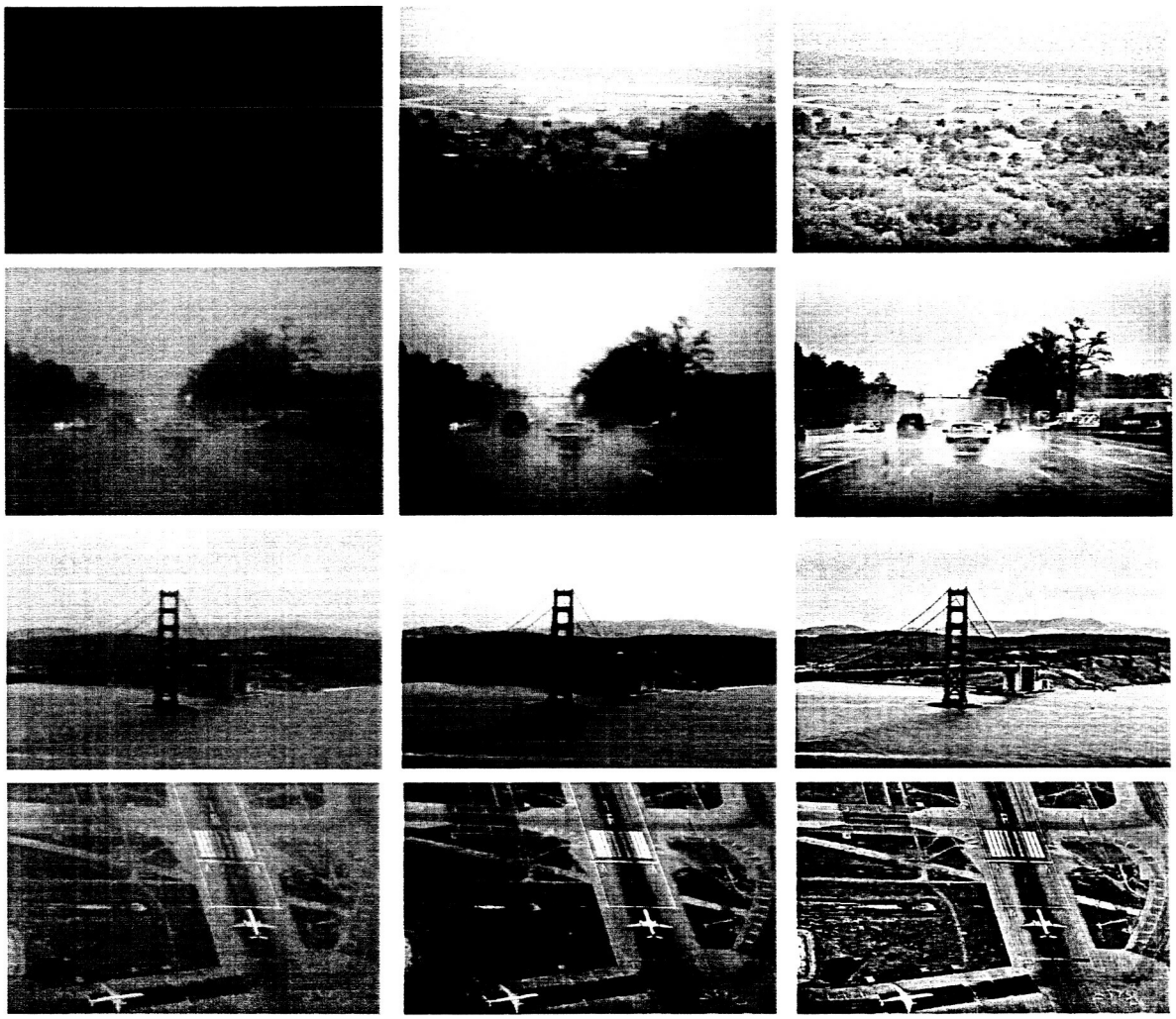


Figure 3. A comparison of automatic visual servo with autolevels. The left column shows the original images; the center column shows the Autolevels enhanced images; and the right column shows the enhancements produced by the Visual Servo.

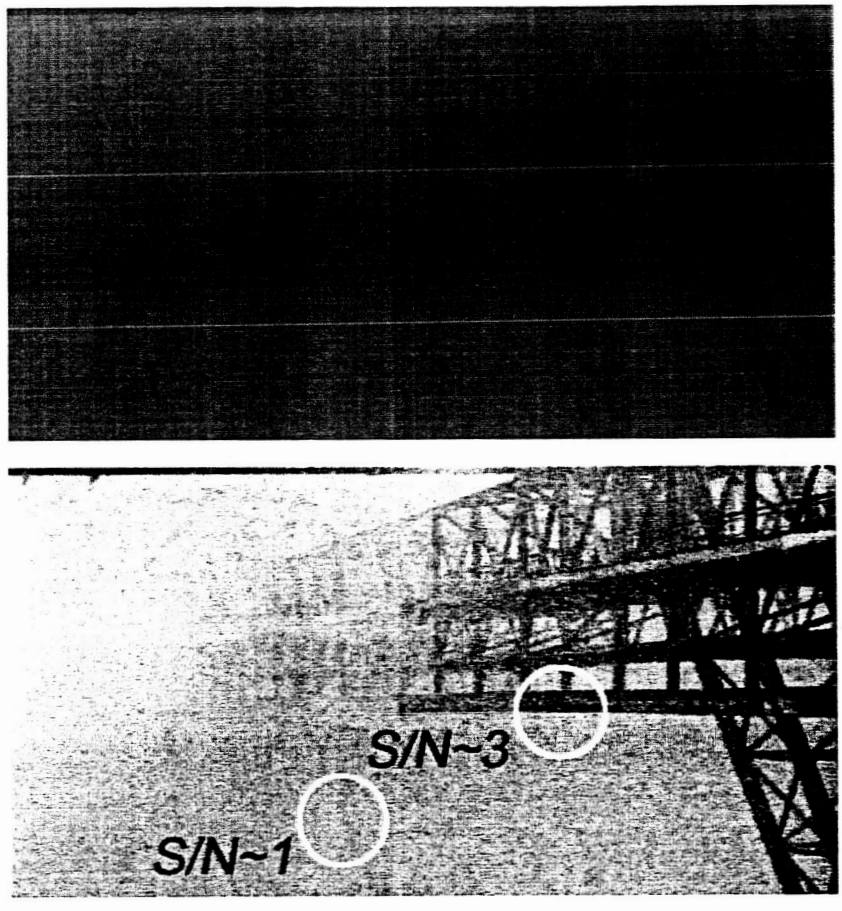


Figure 4: Feature visibility limit of the visual servo

4. THE POST-ENHANCEMENT FEATURE VISIBILITY LIMIT

The dense fog case provides an instructive example for defining feature visibility limits since we encounter post-enhancement opacity immediately and can readily track feature visibility deterioration within short distances. This is shown in Figure 4 where the visibility deterioration occurs for a feature signal-difference approximately equal to the root-mean-square (RMS) noise in the post enhancement image. For slightly closer distances or slightly less turbidity, feature visibility improves rapidly. Even for very low RMS feature signal-difference-to-noise ratios ($\approx$ 3, the visibility is remarkably good as shown in Figure 4). This result appears consistently in a number of other highly turbid test images, so we conclude that post-enhancement feature visibility can be defined by this very simple figure of merit. Unlike more complex figures of merit which must account for both feature signal level as well as feature signal difference, the post enhancement image domain is governed by this noise limit alone. For sensors with lower noise, there will be a consequent improvement in visibility, and ultimately the visibility limit should be set by the signal photon noise, or other scene noises (such as random variations in scattering particle densities) for extremely high sensitivity sensors. This latter case assumes that digitization is done at sufficiently high bit levels so that quantization noise is made lower than any scene noise sources.

5. CONCLUSIONS

An outgrowth of previous developments in non-linear image enhancement was the realization that further improvements in visual contrast and sharpness were needed and that the case of turbid imaging needed to be addressed especially as a significant case for aviation imaging. For this case, a computation was needed which is automatic so that it can be implemented in real-time hardware for enhanced pilot vision during poor visibility flight conditions. These issues together with new scientific insights gained from retinex image processing, led us to develop the VS which is more comprehensive than the previous passive retinex image processing. A set of measures of visual contrast, lightness and sharpness were defined and tested which serve as the basis of the active measurement and control VS system. The system does use non-linear image enhancement for the servo control modules and therefore does not have performance that is derivable from the usual linear systems analysis. Therefore we have extensively tested the VS of wide ranging types and degrees of image turbidity as well as in general purpose imaging.

The servo computation performs well in all turbid imaging condition short of dense fog, and greatly exceeds the human observer's direct perception in all but the dense fog case. The servo handles all manner of moderate fogs, severe hazes, heavy rain, smoke, and heavy snow conditions quite well. It has been tested on color still images as well as color and FLIR video from aviation sensors in varied flight conditions—in and out of clouds, severe haze, twilight-haze, moderate fogs and smoke among others.

Given the use of non-linear image processing, a major issue is how to define a figure of merit for feature visibility limits. An experimental study of post-enhancement image data revealed that feature visibility was limited solely by a very simple figure of merit compared to those used for unenhanced imagery. This figure of merit is that features are visible post-enhancement as long as the feature/background signal-to-noise ratio is greater than unity. Feature visibility increases rapidly for higher signal-to-noise ratios such that a S/N of ≈ 3 has quite good visibility for example.

While our primary interest for the VS is clarifying aviation imagery during poor visibility flight conditions, the servo performs well on underwater turbid images, and in poor visibility road conditions during driving. Therefore we expect that computation has applications to a variety of turbid imaging conditions where a human observer needs to be augmented visually to improve safety and visual performance.

Acknowledgments

Dr. Rahman's work was supported with the NASA cooperative agreement NCC-1-01030. The Visual Servo research was supported by the Synthetic Vision Sensors element of the NASA Aviation Safety Program.

REFERENCES

1. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing* **6**, pp. 451–462, March 1997.
2. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965–976, July 1997.
3. Z. Rahman, D. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," in *Human Vision and Electronic Imaging VII*, B. E. Rogowitz and T. N. Pappas, eds., pp. 390–401, Proc. SPIE 4662, 2002.
4. D. Jobson, Z. Rahman, and G. A. Woodell, "The statistics of visual representation," in *Visual Information Processing XI*, Z. Rahman, R. A. Schowengerdt, and S. E. Reichenbach, eds., pp. 25–35, Proc. SPIE 4736, 2002.
5. F. O. Huck, C. L. Fales, R. E. Davis, and R. Alter-Gartenberg, "Visual communication with Retinex processing," *Applied Optics* , Apr 2000.
6. F. O. Huck, C. L. Fales, and Z. Rahman, "Information theory of visual communication," *Philosophical Transactions of the Royal Society of London A* **354**, pp. 2193–2248, Oct. 1996.

7. C. L. Fales, F. O. Huck, R. Alter-Gartenberg, and Z. Rahman, "Image gathering and digital restoration," *Philosophical Transactions of the Royal Society of London A* **354**, pp. 2249–2287, Oct. 1996.
8. J. A. McCormick, R. Alter-Gartenberg, and F. O. Huck, "Image gathering and restoration: Information and visual quality," *Journal of the Optical Society of America A* **6**(7), pp. 987–1005, 1989.
9. B. D. Thompson, Z. Rahman, and S. K. Park, "A multi-scale retinex for improved performance in multi-spectral image classification," in *Visual Information Processing IX*, Proc. SPIE 4041, 2000.

Statistics of visual representation

D. J. Jobson, Z. Rahman, and G. A. Woodell,

SPIE International Symposium on AeroSense

Visual Information Processing XI

Proceedings SPIE 4736, pp. 25–35

Orlando, FL (2002) (Invited paper)

The Statistics of Visual Representation

Daniel J. Jobson*, Zia-ur Rahman¤, Glenn A. Woodell*
*NASA Langley Research Center, Hampton, Virginia 23681
¤College of William & Mary, Williamsburg, Virginia 23187

ABSTRACT

The experience of retinex image processing has prompted us to reconsider fundamental aspects of imaging and image processing. Foremost is the idea that a good visual representation requires a non-linear transformation of the recorded (approximately linear) image data. Further, this transformation appears to converge on a specific distribution. Here we investigate the connection between numerical and visual phenomena. Specifically the questions explored are: (1) Is there a well-defined consistent statistical character associated with good visual representations? (2) Does there exist an ideal visual image? And (3) what are its statistical properties?

INTRODUCTION

The process of testing, developing, and extensively using the Multiscale Retinex with Color Restoration[1-3] (MSRCR) algorithm for image enhancement has brought forth several fundamental questions about the visual image. The MSRCR is a non-linear spatial and spectral transform that produces images that have a high degree of visual fidelity to the observed scene. In a previous paper,[4,5] we showed that the image of a scene formed using linear representation does not usually provide a good visual representation compared with the direct viewing of the scene. Given that a non-linear transform appears to be essential to the realization of good visual image rendition, we felt a need to further explore the connection between the numerical and the visual representations, i.e. between the numbers that are the digital image, and the visual image that they represent. With the MSRCR we felt we possessed an effective tool for large-scale experimentation and testing on highly diverse images (Figure 1). We asked questions such as: "Is there a statistical ideal visual image?" and "Do all good visual renderings share a convergent statistical character?" These questions, if answered in the affirmative, yield quantitative insights into visual phenomena and lay a general foundation for new definitions of absolute measures of visual quality, which can be used to automatically assess the quality of arbitrary images. Finally, these statistics point to hypotheses concerning the basic mathematical principles of visual representation, which define the general goal of image enhancement in a concise form.

THE INITIAL HYPOTHESIS AND ITS MODIFICATION

As a starting point, we explored the idea that good visual representations seem to be based upon some combination of high regional visual lightness and contrast. To compute the regional parameters, we divide the image into non-overlapping blocks that are 50×50 pixels. For each block, a mean, I, and a standard deviation, σ_I, are computed. A first approach was to postulate that for visually good rendition the contrast×lightness product should be above a minimum value, with the additional constraint that each component cannot fall below an absolute minimum value (Figure 2). This regional scale is sufficiently granular to capture the visual sense of regional contrast. Both the contrast and the lightness can be measured in terms of the regional parameters. The overall lightness is measured by the image mean, $\mu = \bar{I}$, which is also the ensemble measure for regional lightness. The overall contrast, $\bar{\sigma}_I$, is measured by taking the mean of regional standard deviations, σ_I, and it provides a gross measure of the regional contrast variations. The global standard deviation of the image did not relate, except very weakly, to the overall visual sense of contrast. Image frame sizes ranged from 512×512 to 1024×1024 pixels. The coupling of the constraints of minimum contrast-lightness product with minimum contrast and lightness as separate entities defines the zone in Figure 2 labeled "visual good". Further, this figure suggests that there may exist a contour of much higher contrast-lightness, which can be considered a "visual ideal".

DJJ: **d.j.jobson@larc.nasa.gov**; co-authors: ZR: **zrahman@cs.wm.edu**: GAW: **g.a.woodell@larc.nasa.gov**

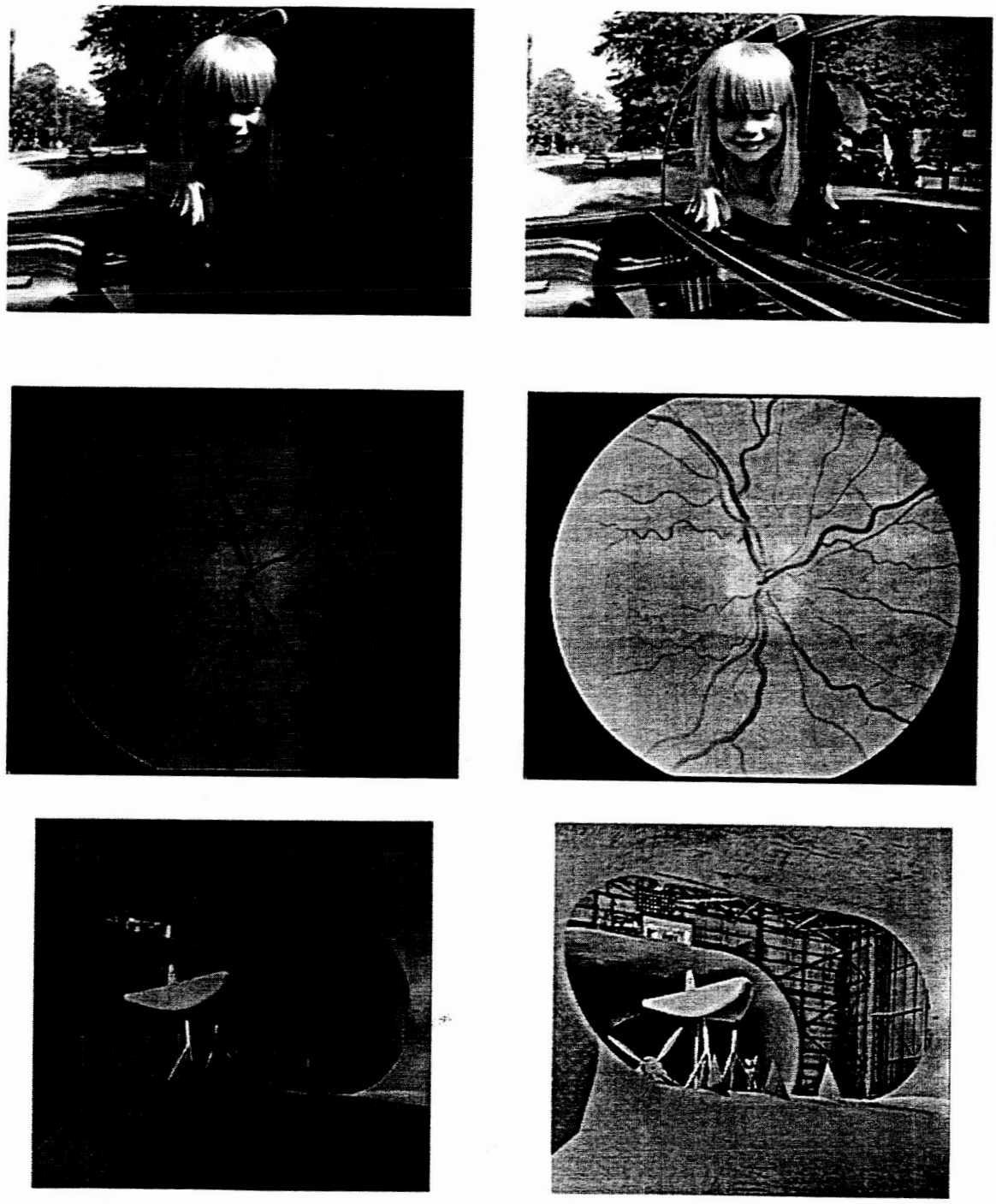


Figure 1. Examples of original and optimized images

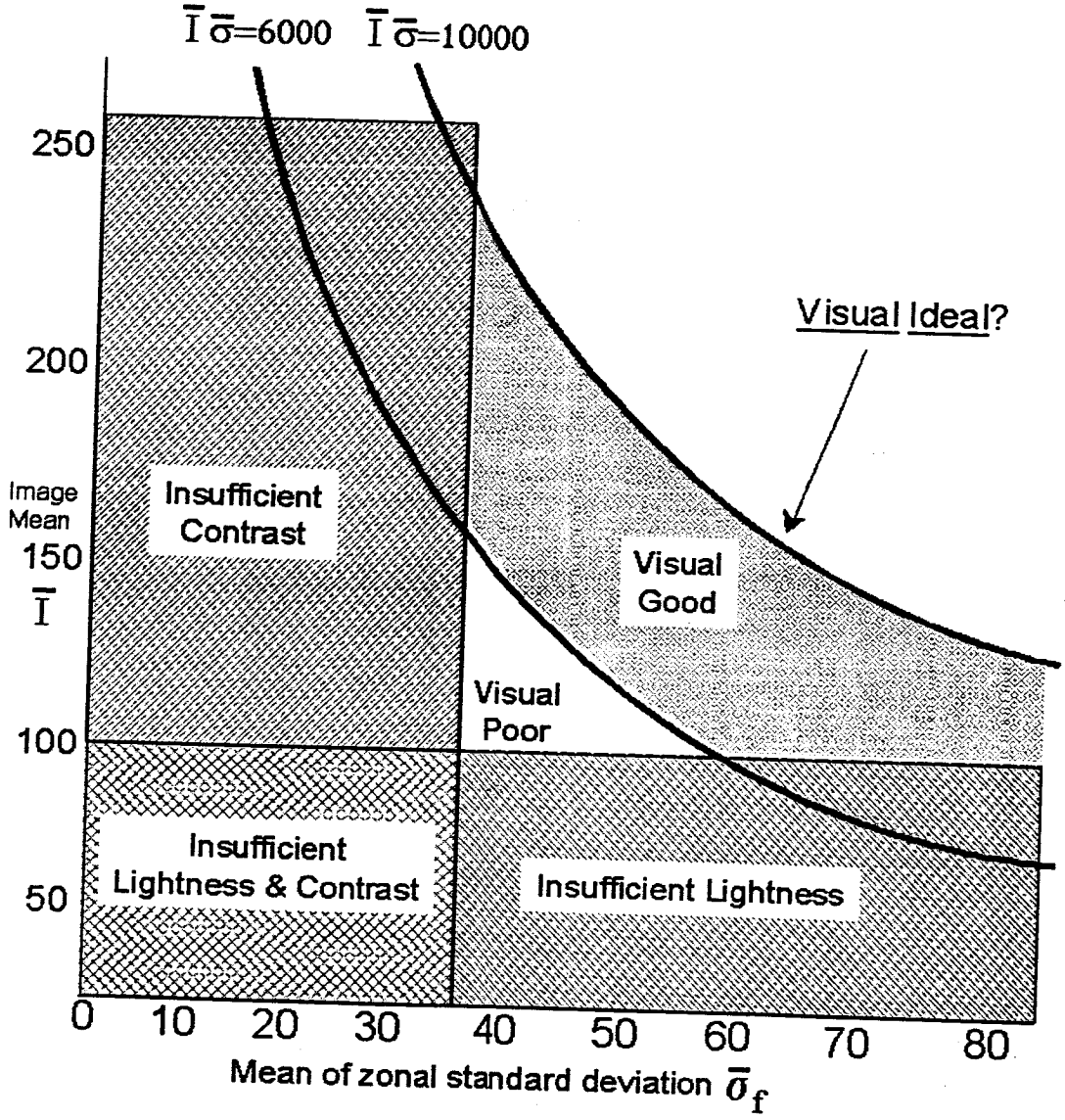


Figure 2. Initial hypothesis; contrast and lightness

To test this hypothesis, we performed some preliminary experiments. The first was to visually optimize a small sample of images using the MSRCR and any other more conventional processing, such as contrast stretch and sharpening. Even this small data set demonstrates that the initial hypothesis is not entirely satisfactory. Though the data exhibit a trend of clustering about a fairly stable mean with quite variable values for the standard deviation, these did not follow any of the particular contours for minimum contrast-lightness product.

For the second experiment we used a larger number of samples (24 images), but otherwise the experiment was identical to the first. The image samples were selected to be as diverse as possible so that the results would be as general as possible. While the MSRCR performs a visually dramatic transformation in most images, the output image can sometimes be further visually optimized, especially by the application of a sharpening filter. This can be expected as a result of vagaries in pre-MSRCR image pre-conditioning, and the blur introduced in the original images by the optics of the image acquisition devices. While the MSRCR is robust with respect to image pre-conditioning, it cannot be completely immune.[6] The post-MSRCR fine-tuning was confined to modest adjustments in brightness and contrast, and sharpening. The results (Figure 3) are shown in stages to make clear the migration (selected points connected by dotted lines) of the data points from the original image data to the MSRCR values then to the visually optimized final destinations. In general, the primary migration is to higher contrast values with relatively smaller increases in lightness. This confirms and quantifies the visual judgment that most images need contrast improvement to be better visual renderings. The visually optimized outputs do converge to a range of approximately 40-80 for global mean of regional standard deviation and global means of 100-200. Again the data do not follow any specific contours for minimum contrast-lightness, but rather appear to be gravitating to a box (Figure 4). So we revise our initial hypothesis accordingly.

There is a sense of increasing visual quality within this box from left to right. To that extent, we can say that the extreme right edge of the box could be regarded as an ideal, but not an ideal that can be realized by all images. Rather it is an ideal that can be approached by some enhanced images. The extreme left of the box is problematic in the following way. The data points here are associated with feature impoverished images—those with large stretches of uniform space—e.g., a small object against a large blank background. Therefore the placement of the left boundary requires a semantic decision and demands a judgment be made about the minimum amount of feature information that an image must contain in order to be considered to be visually good. At the extreme we can certainly agree that the null image cannot be visually good, so at some point we should be able to say an image is intrinsically bad if there is just too much blank space. For a photographer this would correspond to needing to zoom in and have the subject fill more of the blank space in order to have a satisfactory picture. Perhaps in a more informational sense, blank space in images conveys little semantic information except about the relative size of this nullity. This seems intuitively less informative than the world of features, which conveys information about objects and textures.

The issue of convergence for the visually optimized rendition versus original image data requires more experimentation with an even larger more diverse sample. Results exhibit two primary trends summarized schematically in Figure 5(b). Figure 5(a) (for ~100 images) shows the clustering of actual data points. These data support the idea that the visually optimized representations compared to original data do converge in two senses: (1) mean values cluster and do so reasonably tightly around an average of about 165 whereas original image distributions exhibit mean values that scatter rather more evenly across a wider range, and (2) the frame average of regional standard deviations for the visually optimized images all shift to significantly higher values, but do not necessarily converge to any particular value. Further, these same frame averages do shift above a minimum of about 35. Figure 5(b) summarizes these trends for a still larger set of images (~300). So we conclude that these data support the idea that there are distinctive statistical characteristics for good visual representation and that the distinctiveness is sufficiently strong that it can serve as a partial basis for defining new visual measures that automatically assess visual quality. The partial overlap in the two classes in Figure 5 indicates that these two parameters alone are not completely distinguishing. Overall the visually good representation possesses a mean of ~165 and a frame average standard deviation above 35–40. These large samples support the modified hypothesis of Figure 4. A remnant of the initial contour hypothesis (Figure 2) appears possible in Figure 5(a) and is more definite in the actual plots summarized in Figure 5(b). However this statistical tendency is largely overwhelmed by the confines of the box in Figure 4. At the most it appears to be a secondary effect in the statistics of visual representation.

When images are displayed on monitors, their intensity profile is typically modified using the gamma-transformation given by: $I_o(x,y) = [I_i(x,y)]^{1/\gamma}$, where $I_i(x,y)$ is the input value, and $I_o(x,y)$ is the modified value. A value of $\gamma = 1$ is the linear transform. In order to gauge our results against a linear baseline for the original image data, we determined that most digital images are super-linear and should be corrected to approximate linearity by gamma transforming the processed image using $\gamma = 0.63$. While this has negligible impact on standard deviation values, it does adjust the mean downward from about 165 to about 128. The implications of this are discussed next.

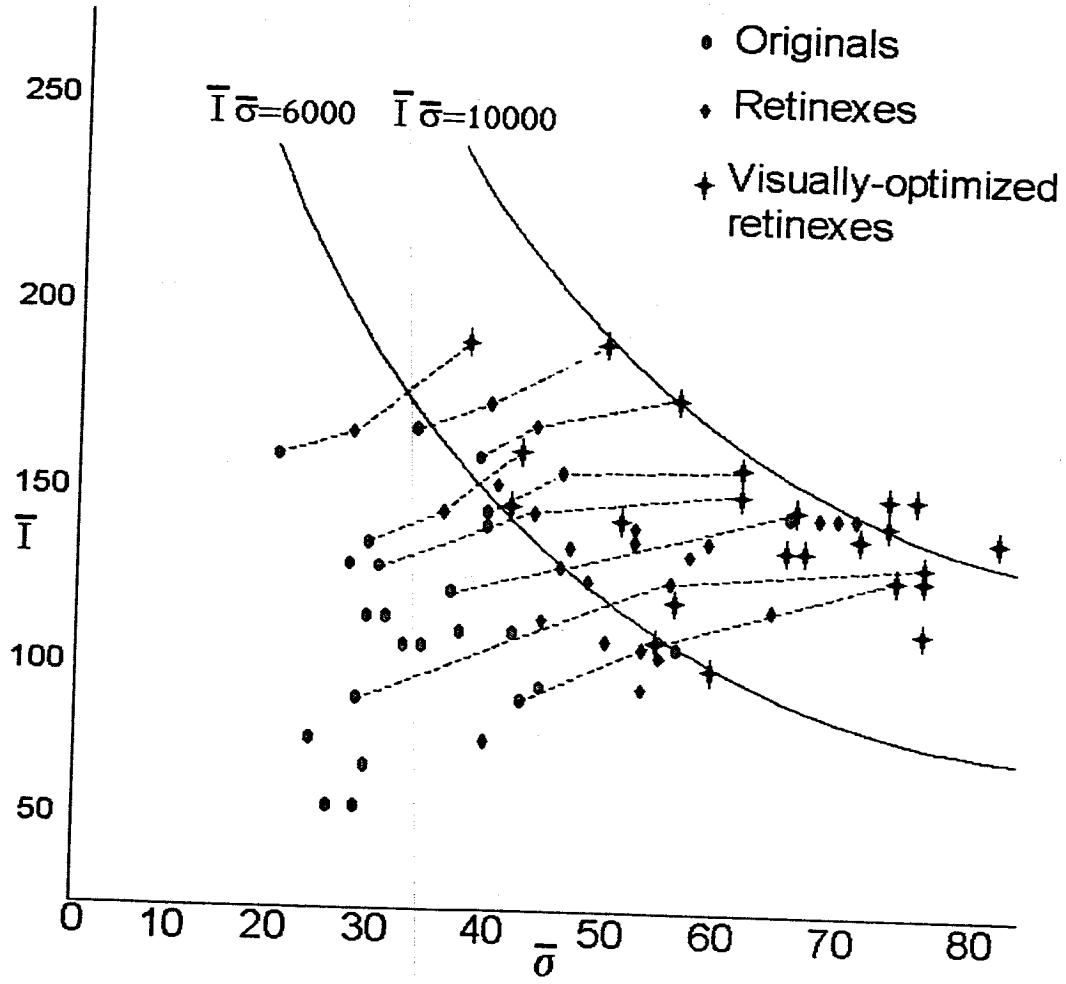


Figure 3. Emerging statistical trends in visual optimization

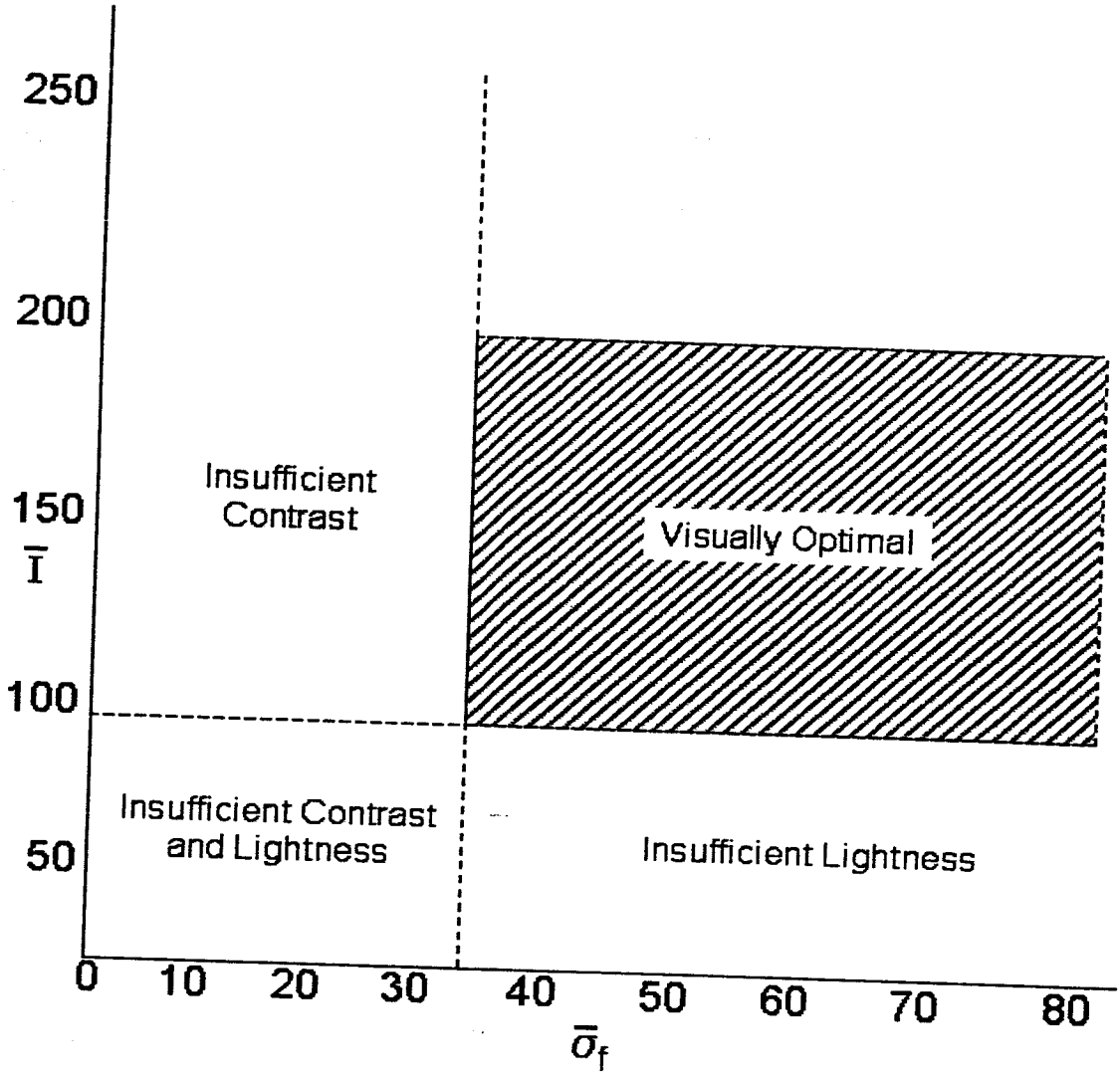


Figure 4. Modified hypothesis

IMPLICATIONS: DEFINING UNDERLYING MATHEMATICAL PRINCIPLES

These data coupled with visual examinations of large numbers of retinex visual optimizations led to the definition of two mathematical principles that are being followed with the observed trends and results. If we view the digital image as a 3-dimensional box (Figure 6), the visually optimal representation appears to jointly satisfy two mathematical conditions for this box. These are:

1. For any spatial scale ranging from near local to near global, visual optimization centers the distributions of regional means near the mid-level of the box (128 for the 8-bit image) and spreads the signal excursions (as quantified by mean of regional standard deviations) out to fill the box as much as possible.

2. Visual optimization spatially minimizes any over- or under-shoots of the box. This is a statement that both clipping to zero and saturation are spatially limited to small zones of infrequent occurrence.

Since the data presented here are all for one spatial scale (the 50×50 pixel region), these two mathematical principles are postulated as working hypotheses concerning the underpinnings of the visual optimization process. Scale changes will not affect the statement about optimization forcing the mean to the midpoint of the dynamic range, but clearly can affect the statement about standard deviation. The statement regarding under- and over-shoots is not scale dependent and appears to be general as long as the original image data prior to optimization is not strongly clipped or saturated over large spatial zones.

These principles suggest that the visually optimal, in more vernacular terms, is centering data on the middle of the box, and spreading the contrast out vertically in the box to the maximum extent possible while minimizing excursions outside the box. This certainly seems to be a pre-conditioning of image signals to most efficiently occupy the box space.

DISCUSSION: THE IDEAL VISUAL REPRESENTATION

The mainstream of the data presented here is associated with optimization to the point of producing a good visual representation. But it is interesting to consider what might be an ideal visual representation. In preceding discussion we noted that feature-impoverished images are debatable cases, and that there seems to be a minimum of feature occurrence necessary just to achieve a "good" visual image. At the opposite extreme, an ideal visual representation fundamentally needs to be feature-rich and the optimization needs to achieve a strong sense of visual contrast approaching that found in the graphics world of illustration. Clearly this is not possible, as already noted, for all images. So only a restricted class of images are even candidates for an optimization that approaches some ideal. In numerical terms, we can see that the mean value of about 165 should not be affected by "good" versus "ideal," but that the "ideal" will exhibit much larger values of contrast (standard deviation) in the range of 60-90. Images, which can be enhanced to this level, are ones for which there is a high degree of reflectance diversity in the scene, rich feature densities, and successful retinex dynamic range compression for scenes, which have strong lighting variations. An extreme case, which does not have the visual sense of being "ideal" is the printed text image. While text images do have high standard deviations (~90), they do not represent natural scenes, and can be compressed to binary data (not needing 8 or morebits). Further we think that the act of reading is far more of a local raster scanning process than the more global visual sense of comprehending pictures. The visual judgment applied to pictures is therefore not likely to be involved in the reading of text.

The "ideal" should however be associated with a near-perfect sense of clarity and sharp features so sharpness is important component of "ideal" which we do not specifically address here. We did however make frequent use of post-retinex sharpening to reach the visual optimizations. These considerations of "ideal" cannot be related to aesthetics, where often the diffuse or impressionistic or murky are the most beautiful and may be "ideal" in that strictly aesthetic sense.

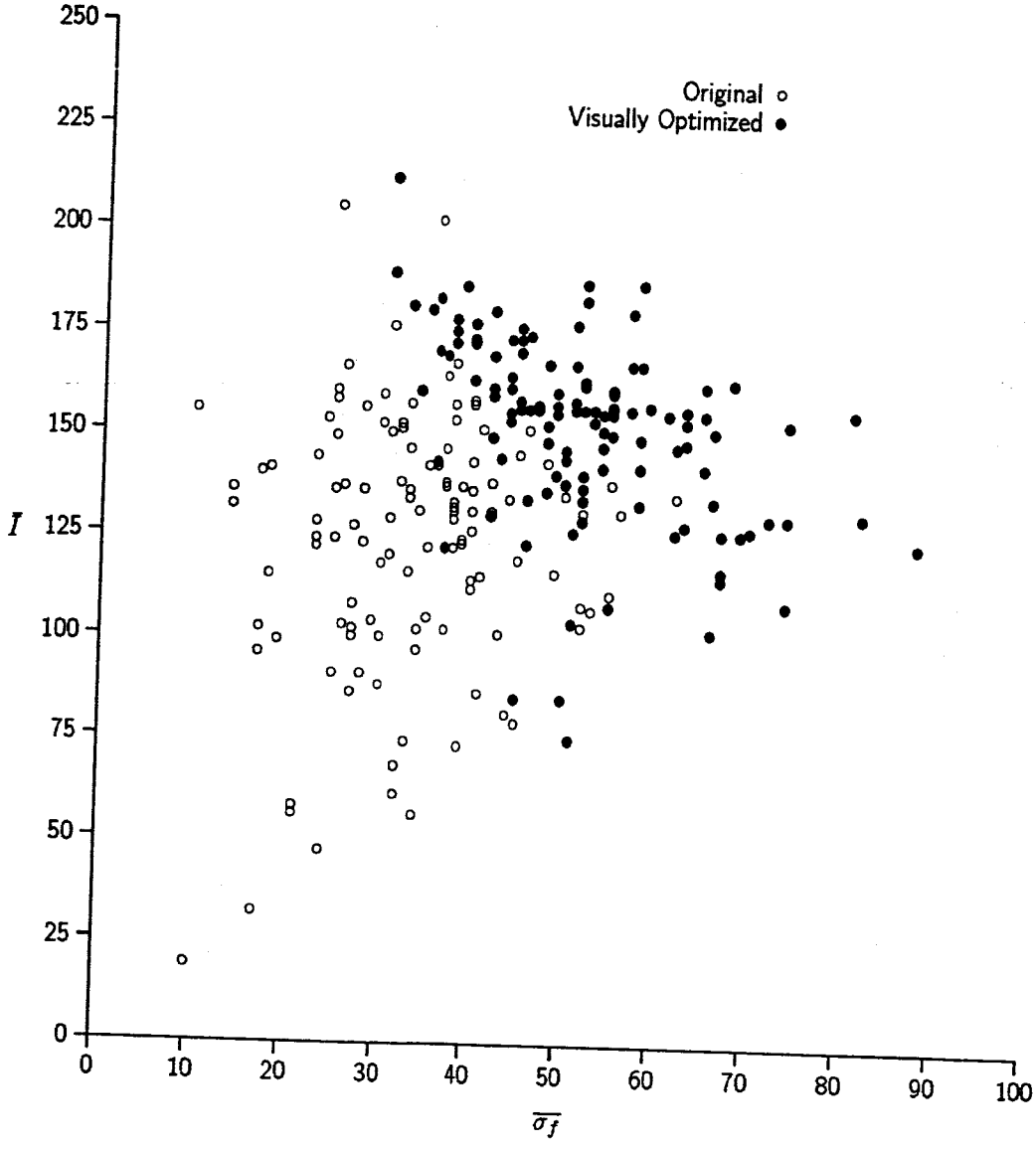


Figure 5(a). Large sample of original and optimized images

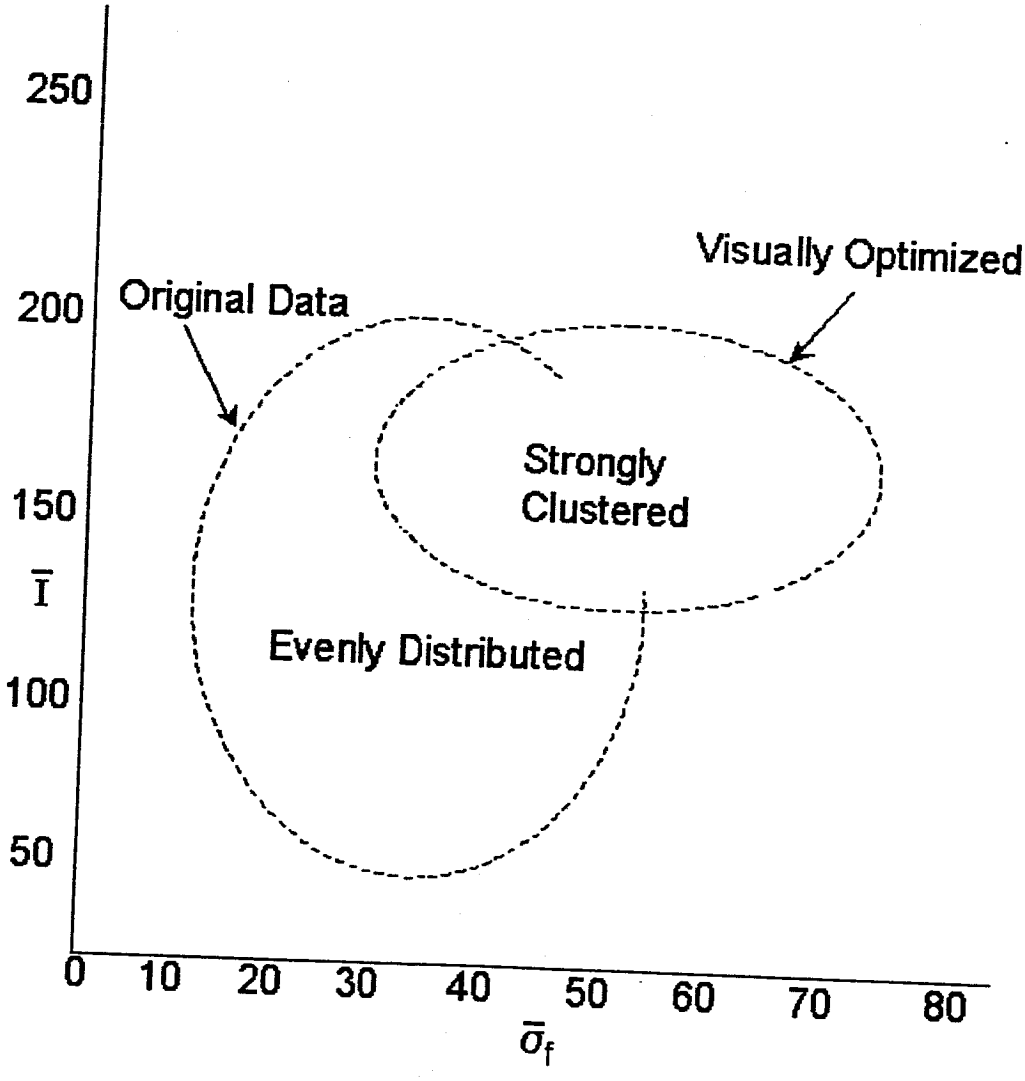


Figure 5(b). Large sample - overall trends in optimization

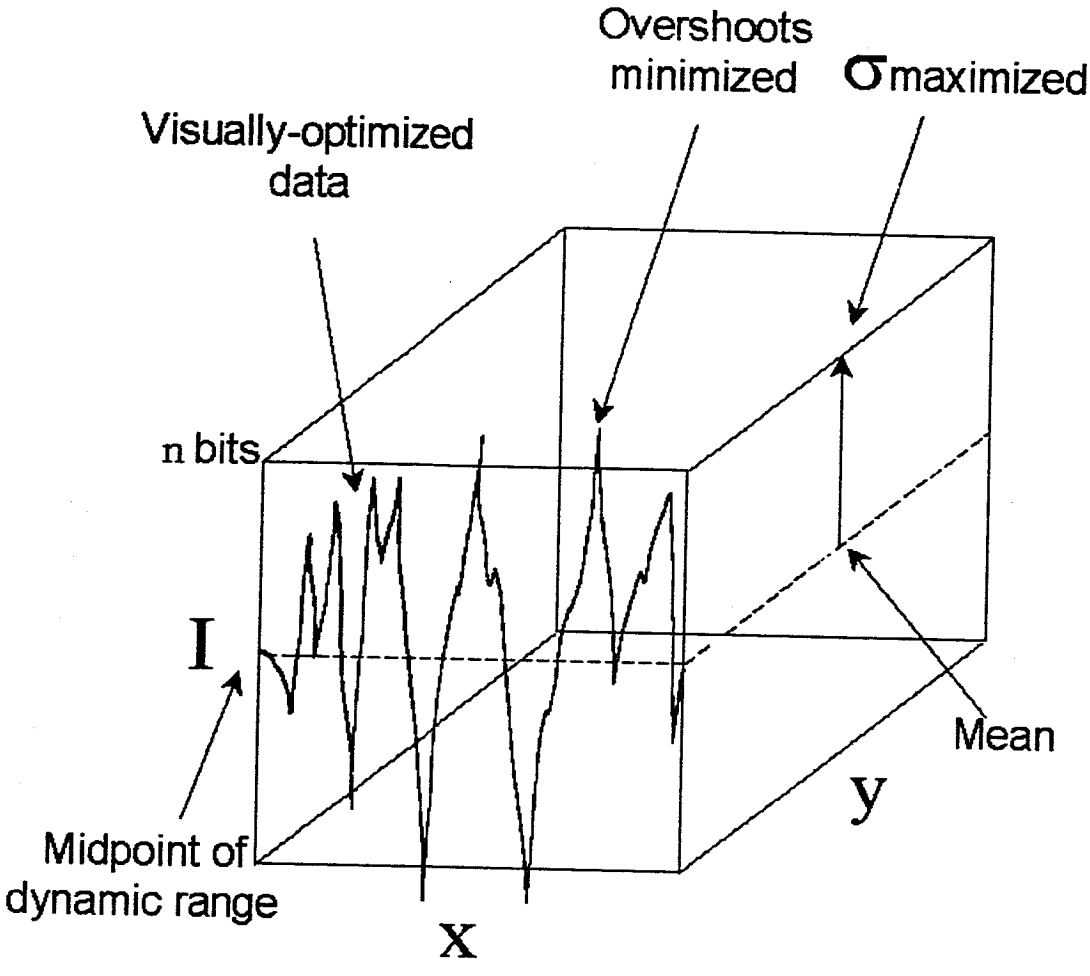


Figure 6. Underlying mathematical principles of optimization

CONCLUSIONS

Guided by the extensive experience of enhancing images using retinex methods, we find that good visual representations require a non-linear spatial and spectral transform of raw digital image data, which results in consistent statistical trends. These trends provide a new quantitative understanding of the goals of image processing for visual rendition and a partial foundation for constructing visual measures for automatically assessing the quality of visual representation. In general, visually optimized images are more tightly clustered about a single mean value and have much higher standard deviations. Further the results support the idea that visual optimization centers the data mean on the mid-point of the image dynamic range and spreads the signal excursions out across the dynamic range to a maximal extent while at the same time limiting any over- and under-shoots spatially. This overall trend relates to most efficiently occupying the data space with the actual image data. In general visually optimized images are improved in terms of both regional lightness and contrast with the latter being the most strongly affected.

REFERENCES

1. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A Multi-Scale Retinex For Bridging the Gap Between Color Images and the Human Observation of Scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965-976, July 1997.

2. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and Performance of a Center/Surround Retinex," *IEEE Transactions on Image Processing* **6**, pp. 451-462, March 1997.

3. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multiscale Retinex for Color Rendition and Dynamic Range Compression," in *Applications of Digital Image Processing XIX*, A. G. Tescher, ed., Proc. SPIE 2847, 1997.

4. D. J. Jobson, Z. Rahman, and G. A. Woodell, ``Spatial Aspect of Color and Scientific Implications of Retinex Image Processing," in *Visual Information Processing X*, S. K. Park, Z. Rahman, and R. A. Schowengerdt, eds., pp. 117-128, Proc. SPIE 4388, 2001.

5. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," in *Human Vision and Electronic Imaging VII*, B. E. Rogowitz and T. N. Pappas, eds., Proc. SPIE 4662, 2002.

6. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Resiliency of the Multiscale Retinex Image Enhancement Algorithm," in *Proceedings of the IS&T Sixth Annual Color Conference: Color Science, Systems, and Applications*, pp. 129-134, IS&T, 1998.

Multi-sensor fusion and enhancement using the Retinex image enhancement algorithm

Z. Rahman, D. J. Jobson, G. A. Woodell, and G. D. Hines,

SPIE International Symposium on AeroSense

Visual Information Processing XI

Proceedings SPIE 4736, pp. 36-44

Orlando, FL (2002)

Multi-sensor fusion and enhancement using the Retinex image enhancement algorithm

Zia-ur Rahman[†], Daniel J. Jobson[‡], Glenn A. Woodell[‡], Glenn D. Hines[‡]

[†]College of William & Mary, Department of Computer Science, Williamsburg, VA 23187.
[‡]NASA Langley Research Center, Hampton, Virginia 23681.

ABSTRACT

A new approach to sensor fusion and enhancement is presented. The retinex image enhancement algorithm is used to jointly enhance and fuse data from long wave infrared, short wave infrared and visible wavelength sensors. This joint optimization results in fused data which contains more information than any of the individual data streams. This is especially true in turbid weather conditions, where the long wave infrared sensor would conventionally be the only source of usable information. However, the retinex algorithm can be used to pull out the details from the other data streams as well, resulting in greater overall information. The fusion uses the multiscale nature of the algorithm to both enhance and weight the contributions of the different data streams forming a single output data stream.

1. INTRODUCTION

The Multiscale Retinex (MSR) was developed as a general-purpose image enhancement algorithm that provided simultaneous dynamic range compression and color constancy. Aside from the obvious applications to the correction of color and grayscale digital images, this algorithm can also be applied to enhance infrared and near-infrared spectral bands. In addition, due to the inherent multiscale structure, it can be used to emphasize different features in different spectral bands, yet provide a composite image that contains more information than any individual spectral band. It is the combination of these properties that led us to investigate the MSR as a possible fusion engine. In this paper we will present the results of two different approaches for using the MSR as a fusion engine. The primary driving force for developing these fusion strategies was to provide an image that could be used to "see" in poor visibility conditions such as dusk, and fog.

The Retinex (retina+cortex) was developed as a model of the human color vision system by Edwin Land.[1–3] The MSR was developed by Jobson, et al[4,5] to address the problems due to changing lighting and atmospheric conditions that were inherent to the process of acquiring images of Earth from space. The MSR provides the dynamic range compression, color constancy and sharpening that are required to alleviate these problems. It has been shown that applying the MSR to remotely-sensed images improves the overall classification accuracy.[6,7] In addition, the MSR processed images allow a greater number of features to be detected than would be possible with un-enhanced data. Both of these results are also of importance in the context of sensor fusion because they provide more accurate registration points between the data streams that need to be fused.

The MSR belongs to the class of center-surround operators and can be succinctly expressed as

$$\mathcal{R}_i(x_1,x_2) = \sum_{k=1}^{K} \mathcal{W}_k\big(\log[\mathcal{I}_i(x_1,x_2)] - \log[\mathcal{I}_i(x_1,x_2) * \mathcal{S}_k(x_1,x_2)]\big), \quad i = 1,\ldots,\mathcal{N}, \tag{1}$$

where $\mathcal{R}_i$ and $\mathcal{I}_i$ are, respectively, the ith spectral band of the MSR output and the input, $\mathcal{W}_k$ and $\mathcal{S}_k$ are, respectively, the weight and the surround function associated with the kth scale, K is the number of scales, and $\mathcal{N}$ is the number of spectral bands. The kth surround function is given by

$$\mathcal{S}_k(x_1,x_2) = \alpha \exp[-(x_1^2 + y_1^2)/\sigma_k^2],$$

ZR: zrahman@cs.wm.edu; DJJ: d.j.jobson@larc.nasa.gov; GAW: g.a.woodell@larc.nasa.gov; GDH: g.d.hines@larc.nasa.gov

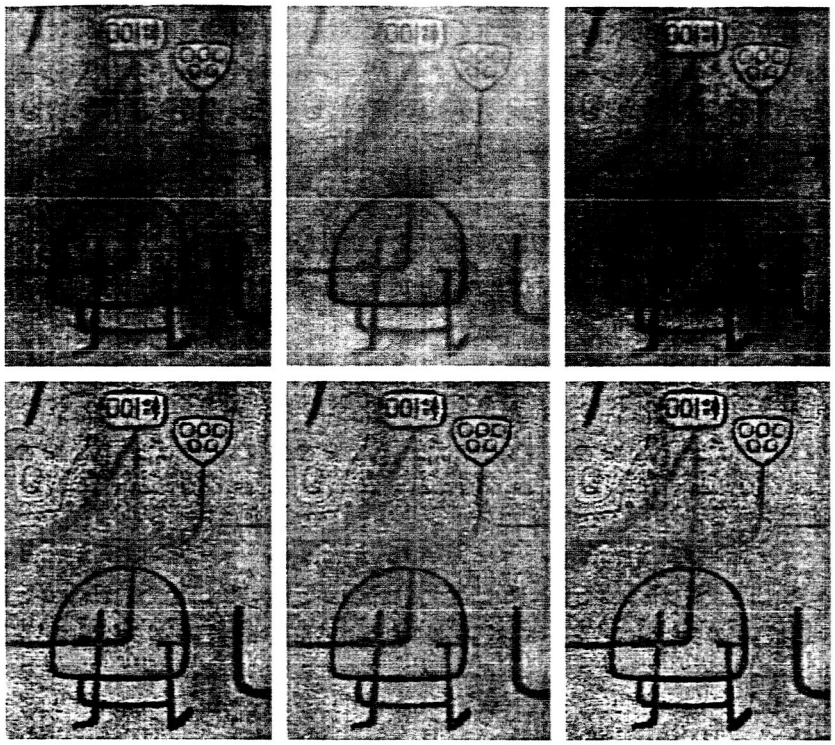


Figure 1: Example of color constancy and feature sharpening.

where σ_k is the scale factor that controls the width of the kth surround function, and $\alpha = \left(\sum_{x_1,x_2} \mathcal{S}_k(x_1,x_2)\right)^{-1}$ is the normalization factor. The color constancy of the MSR is a direct result of the center-surround nature of the algorithm. Equation 1 can be rewritten as:

$$\mathcal{R}_i(x_1,x_2) = \sum_{k=1}^{K} \mathcal{W}_k \left(\frac{\log[\mathcal{I}_i(x_1,x_2)]}{\log[\mathcal{I}_i(x_1,x_2) * \mathcal{S}_k(x_1,x_2)]} \right), \quad i = 1,\ldots,\mathcal{N}. \tag{2}$$

The intensity values in each spectral band at location (x_1,x_2) can be written as the product of two components: $\rho(x_1,x_2)$, the reflectance component, which represents the light reflected or emitted from all the objects in the scene, and $\iota(x_1,x_2)$ which represents the illumination component. That is,

$$\mathcal{I}_i(x_1,x_2) = \iota_i(x_1,x_2)\rho(x_1,x_2).$$

Since the illumination component varies very slowly across the scene, $\iota(x_1,x_2) \approx \mathcal{I}_o$, and $\mathcal{I}(x_1,x_2) \approx \mathcal{I}_o\rho(x_1,x_2)$. Thus,

$$\mathcal{R}(x_1,x_2) = \log\left(\frac{\mathcal{I}_o\rho(x_1,x_2)}{\mathcal{I}_o\rho(x_1,x_2) * \mathcal{S}(x_1,x_2)}\right) = \log\left(\frac{\rho(x_1,x_2)}{\rho(x_1,x_2) * \mathcal{S}(x_1,x_2)}\right),$$

which is independent of the illuminant. An example of this is shown in Figure 1 where the top row shows an image acquired under three different illumination conditions* and the bottom row shows the MSR outputs. Aside from the color constancy, the features in each of the processed images are much sharper, providing the common registration points that are necessary for image fusion.

*Unfortunately, the illumination effects are not readily visible in the grayscale rendition of the color images.

2. SENSOR FUSION

The particular problem that we are trying to address is to fuse data from three different imaging sources: long-wave infrared (LWIR), short-wave infrared (SWIR), and visible color (RGB). In addition to imaging at a different wavelength, these sensors also differ in resolution, and field of view (FOV). Thus, the first processing step is to perform a multi-modal registration of the images from the three different sensors. Table 1 shows the characteristics of the various sensors.[8] Image registration can generally be defined as a mapping between two

	LWIR	SWIR	CCD
Pixel Resolution (nominal)	$320H \times 240V$	$320H \times 240V$	$542H \times 497V$ (RGB)
Optics FOV	$39°H \times 29°V$	$34°H \times 25°V$	$34°H \times 25°V$
Detector readout frame rate	60Hz	60Hz (typical)	30Hz (interlaced)

Table 1: Selected sensor characteristics

or more images both spatially and with respect to intensity. Mathematically $I_2 = g(I_1(f(x,y)))$ where I_1 and I_2 are two-dimensional images (indexed by x and y), f is a transformation of spatial coordinates and g is a one-dimensional intensity or radiometric transformation.[9]

For the spatial transformation our primary complication is that there are no fiducial markers within the images. The cameras are, however, bore-sighted so they all have a common center of alignment. Thus we simply calculate an affine transformation performing the operations of scaling, translation and rotation globally on the images matching them to a common grid. A radiometric transformation could also be performed at this stage since grayscale characteristics, in particular polarity reversal between visible and IR images, may differ locally. Instead, as will be discussed in the next section, we utilize the MSR transformation. Further details regarding the rectification process are not discussed in this paper. We will assume, however, that appropriate rectification has been performed on each of the data streams with the resultant registration of features and correction of FOV and pixel resolution.

2.1. The first approach

Our first approach is relatively straightforward. Using the MSR as the enhancement engine, each of the data streams is individually enhanced to bring out the desired features. The MSR has previously been optimized for the RGB images but not for the LWIR, and SWIR images. So the first task was to analyze the data and determine the optimal parameters that were needed to achieve the desired enhancement in the LWIR and SWIR. This was done experimentally by applying the MSR given in Equation 1 to the LWIR and SWIR images, and observing the output images. The optimization, in this case, is simply the observer's judgment that a given image has more "useful" information than any others observed.

Once the optimized, MSR enhanced images were obtained, the RGB image was converted to grayscale (GS) by using the standard RGB to GS transform:

$$GS = 0.299R + 0.587G + 0.114B \tag{3}$$

where R, G, and B are the red, green, and blue color components of the RGB image respectively. Consistent with the contrast sensitivity of the human visual system,[10] the green channel is weighted most heavily when forming the GS image. After the conversion from RGB to GS, we have three remaining bands of data—GS, LWIR, SWIR—that we need to fuse. A quick way to do this is to use Equation 3 and use the three data bands in place of the R, G, and B components. The obvious question that arises is which data band should be substituted for which color component. Using the contrast sensitivity analogy, the data band which replaces the G component will have the most weight, so the fused image will be most sensitive to changes in this component. It is difficult to determine beforehand, which of the three bands contains the most relevant information and hence should be weighted most heavily. However, since the thrust for the sensor fusion is to provide better visibility in poor visibility conditions, we cannot, in general, rely on the GS band to provide the primary information. This task will typically fall to the LWIR band, especially under foggy conditions. The GS band does have the potential to

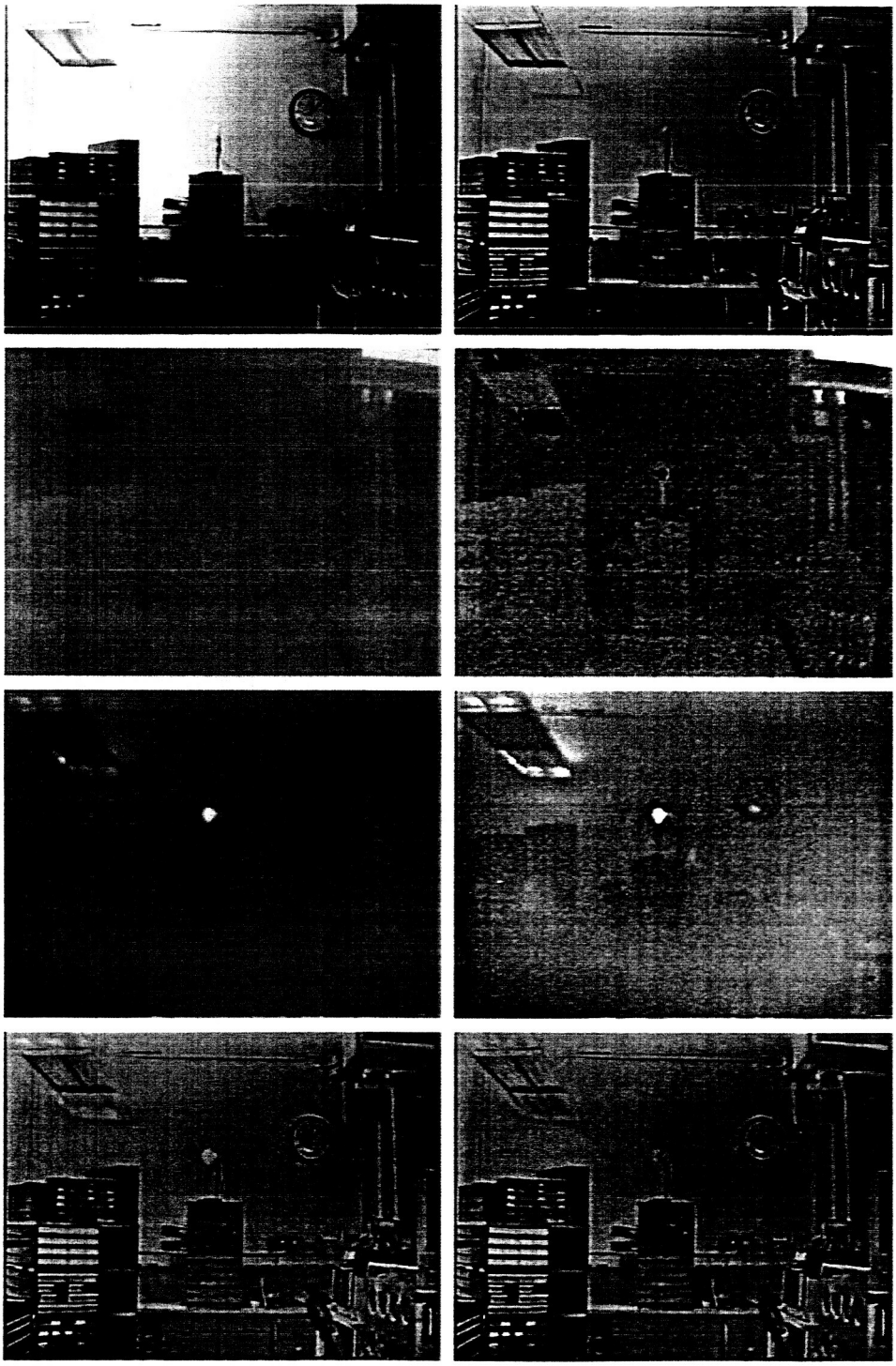


Figure 2. Original (left column) and MSR enhanced images (right column) are shown for RGB (first row), LWIR (second row), and SWIR (third row) images. The bottom row shows the composite images obtained by giving the strongest weight to the GS band: the left column shows the image obtained when the LWIR band is given the next heaviest weight and the right column shows the image when the SWIR band is given the next heaviest weight.

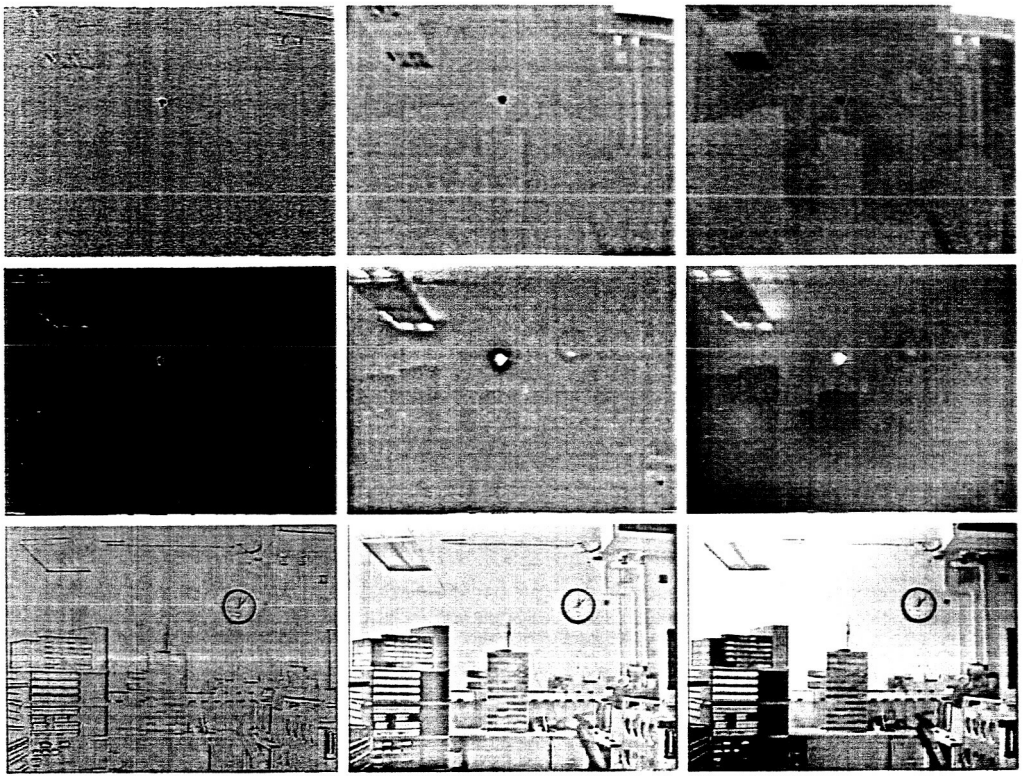


Figure 3. First column: MSR with $\sigma = 5$; middle column: MSR with $\sigma = 20$; last column: MSR with $\sigma = 240$. Top row: LWIR; middle row: SWIR; bottom row: GS. The narrow surround, $\sigma = 5$, acts as a high-pass filter, capturing all the fine details in the image but at a severe loss of tonal information. The wide surround, $\sigma = 240$, captures all the fine tonal information but at the cost of dynamic range. The medium surround, $\sigma = 20$, captures both dynamic range and tonal information.

provide the relevant information under low light conditions, such as dusk. The fused image, I_F, is then obtained by:

$$I_F = 0.299\text{SWIR} + 0.587\text{LWIR} + 0.114\text{GS}, \tag{4}$$

though other combinations are also possible.

Figure 2 shows the original and MSR enhanced LWIR, SWIR, and GS images, as well as the fused data I_F for images that were taken in a laboratory with lights dimmed to simulate lighting conditions similar to those at dusk. In this situation, the original GS has the most information but suffers from lack of information in certain regions in the image such as the top-left where the light from the fluorescent source saturates the image. Both the SWIR, and the LWIR images have information in this region. The MSR enhanced data provides more visual information in each of the bands—see the first three rows in the left column in Figure 2. However, the data from the SWIR and LWIR sensors is difficult to interpret on its own. The bottom row of Figure 2 shows two fused images obtained by:

$$I_{F_1} = 0.299\text{LWIR} + 0.587\text{GS} + 0.114\text{SWIR} \tag{5}$$

$$I_{F_2} = 0.299\text{SWIR} + 0.587\text{GS} + 0.114\text{LWIR} \tag{6}$$

where I_{F_1} is the image shown in the bottom-left of Figure 2 and I_{F_2} is show in bottom-right. Note in particular the additional information in the fused image around the two light sources: this information was missing in the GS image. Both of the fused images provide more information than any of the three data bands can individually.

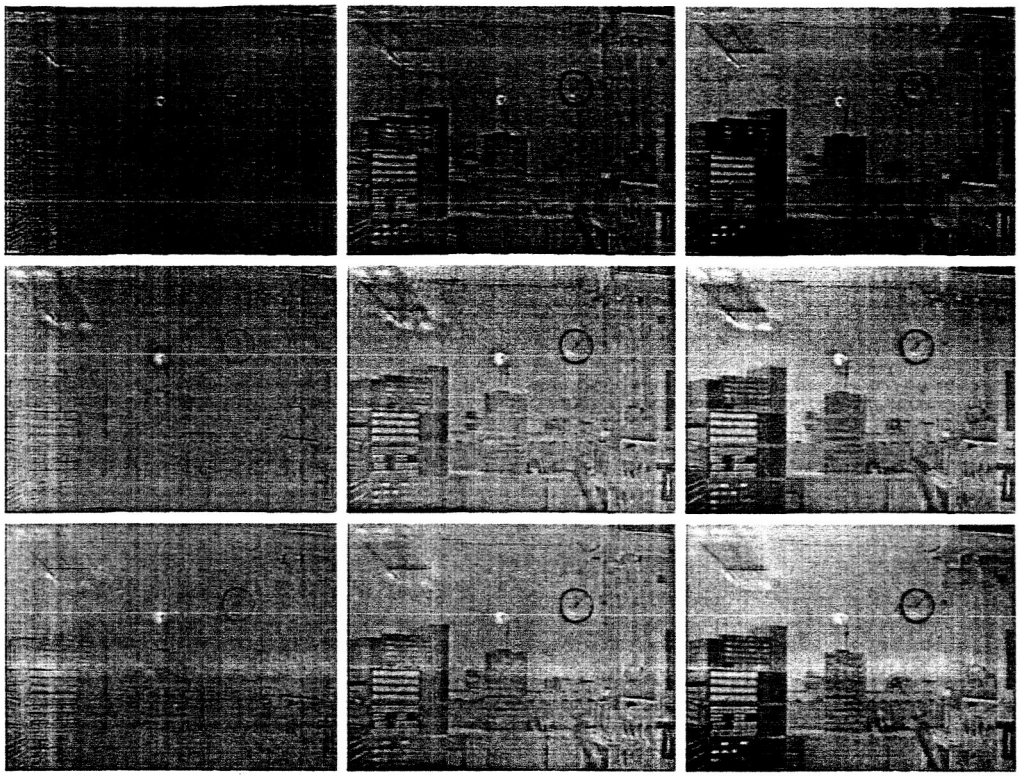


Figure 4. The fused images: $(\sigma_L, \sigma_S, \sigma_G)$ represents the scale values with which the LWIR, SWIR, and GS bands are processed for the SSR. Top row: $(5,5,5)$, $(5,5,20)$, $(5,5,240)$. Middle row: $(5,20,5)$, $(5,20,20)$, $(5,20,240)$. Bottom row: $(5,240,5)$, $(5,240,20)$, $(5,240,240)$.

.

Additionally, the compactness of the fused data, one image instead of three, also makes it more practical for the pilot in terms of ease of viewing and interpretation.

2.2. The second approach

Though the results from the first approach are quite encouraging, we wanted to test a second approach that uses the MSR in a different way. The first approach primarily relied on the dynamic range compression capabilities of the MSR to pull out features that were "difficult" to see in the existing lighting conditions, and then used Equations 5, and 6 to weight the contributions of the different data bands. In our second approach, we use a different mechanism. The MSR as shown in Equation 1 can be rewritten as

$$\mathcal{R}_i(x_1, x_2) = \sum_{k=1}^{K} \mathrm{SSR}_k(\mathcal{I}_i)(x_1, x_2) \quad i = 1, \ldots, \mathcal{N}, \tag{7}$$

where $\mathrm{SSR}_k(\mathcal{I}_i)$ is the single scale retinex (SSR) with the kth scale factor acting on the ith input band $\mathcal{I}_i$. As stated above, the magnitude of the kth scale factor, σ_k, controls the width of the surround function. This, in turn, determines the type and amount of "enhancement" that takes place: a large σ_k retains more color information at the cost of spatial detail, and a small σ_k enhances the spatial detail at the cost of color information. The MSR approach was developed to balance this tradeoff between color information and spatial detail by merging representations that brought out detail and color. Hence small details in the scene can be obtained by using using the MSR with $\mathcal{N} = 1$ and $K = 1$, and a small value of σ_k. This tradeoff between color information and spatial detail is shown in Figure 3.

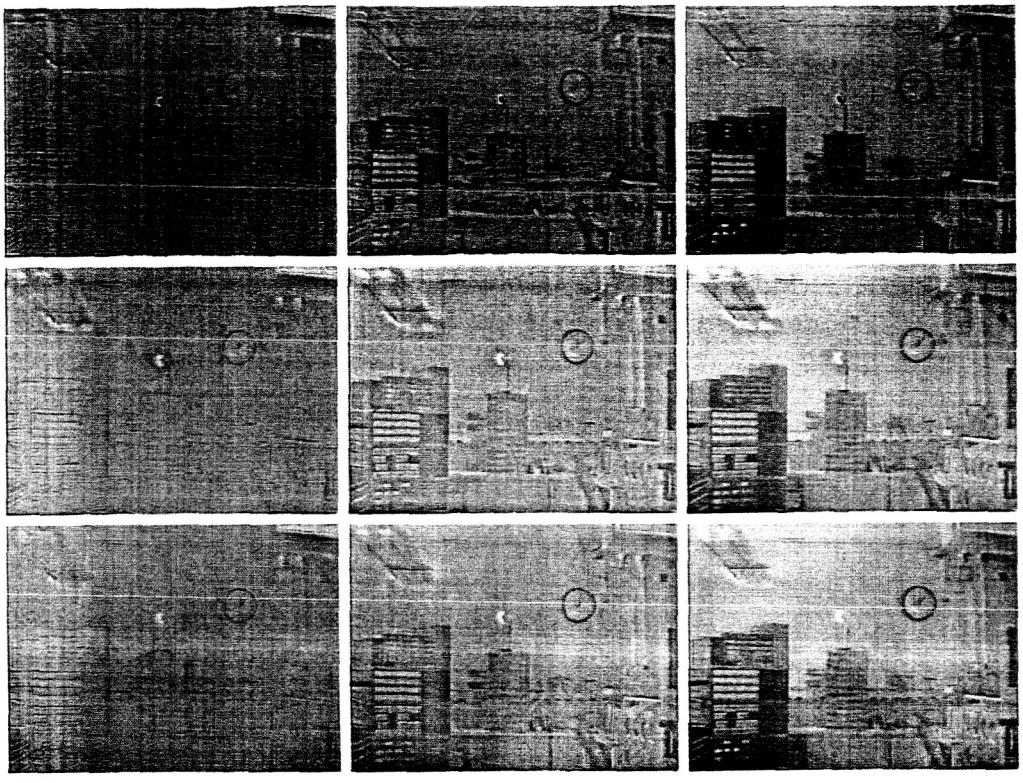


Figure 5. Top row: (20, 5, 5), (20, 5, 20), (20, 5, 240). Middle row: (20, 20, 5), (20, 20, 20), (20, 20, 240). Bottom row: (20, 240, 5), (20, 240, 20), (20, 240, 240).

Using this formulation, we used the SSR to pull out different details from the different data bands before fusing the data together. The fused images can be obtained by averaging the results of any three of these nine renditions that are shown in Figure 3, with the condition that one image from each of the SWIR, LWIR, and GS bands must be used. So, for example, the fused image can be obtained by

$$I_F = \big(\mathrm{SSR}_1(\mathrm{LWIR}) + \mathrm{SSR}_2(\mathrm{SWIR}) + \mathrm{SSR}_3(\mathrm{GS})\big)/3 \tag{8}$$

where SSR_1 was generated with $\sigma_1 = 5$, SSR_2 was generated with $\sigma_2 = 20$, and SSR_3 was generated with $\sigma_1 = 80$. However, with the three bands each processed at three different scales, there is a total of 27 combinations that are possible: Figures 4–6 show all of these various combinations. As is readily evident, some of these combinations are not very useful when used for visual interpretation. However, the fine details in the combinations using the small scales maybe ideal for projection on a heads up display (HUD). The best visual result is obtained when $\sigma_L = 240, \sigma_S = 20$, and $\sigma_G = 240$ (Figure 6). This makes intuitive sense as well because in this case, the GS scale image provides the tonal information for the rendition, and the LWIR, and SWIR images provide the complementary features that are not visible in the GS band. A closer look at Figure 3 shows the complementary features in the three data bands. The amount of visual information that can be obtained from the fused image, as in the first experiment, is considerably more than the amount of visual information available in the individual data bands or in their SSR enhanced versions. This approach has the additional merit in that it allows extraction of fine features from the various data bands, hence making the task of seeing objects, large *and* small, in poor visibility conditions considerably easier. Interestingly, the best result in the second approach is obtained when the second approach is, in essence, a single scale version of the first approach.

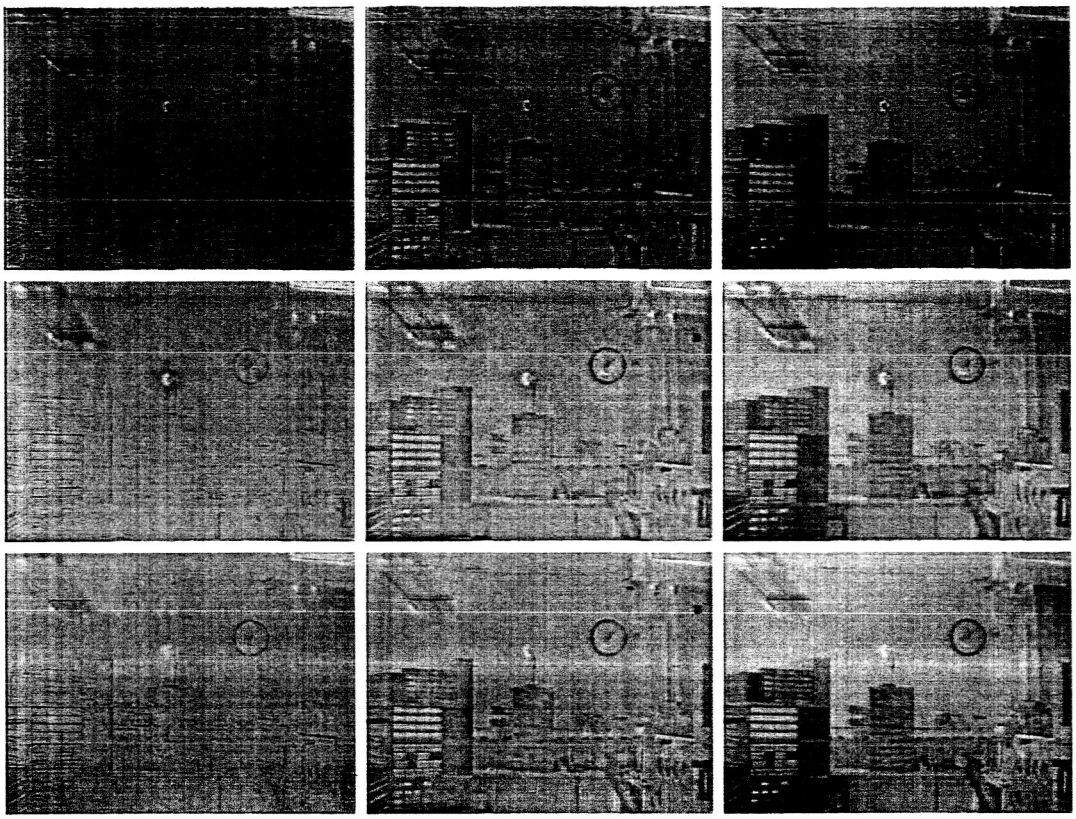


Figure 6. Top row: (240, 5, 5), (240, 5, 20), (240, 5, 240). Middle row: (240, 20, 5), (240, 20, 20), (240, 20, 240). Bottom row: (240, 240, 5), (240, 240, 20), (240, 240, 240).

3. CONCLUSIONS

The task of landing an aircraft in poor visibility conditions is one of the most challenging that a pilot faces. Any technique which allows the pilot to alleviate the problems due to poor visibility will result in safer flights. The two methods we presented in this paper attempt to accomplish this task. Using data from three different sources, long-wave infrared, short-wave infrared, and color, we provide as an end-product, a single fused image that contains more visual information than do any of the data streams individually. The MSR and SSR algorithms are used to process the original data streams, providing enhanced data streams which contain data that is sharper, and has better contrast: i.e., data that provides better visibility.

Two different approaches are presented, both for the same end-goal: better visibility in poor visibility conditions. The techniques differ in some basic ways such as how the different data streams are fused together and which data stream contains more useful features, but each fused data stream has more, and better, information than the original data streams. In addition, the fused data stream allows the data from different sensors to be presented compactly as a single image. This make the pilot's task simpler in that a single image needs to be interpreted rather than three separate ones. If the pilot were required to interpret several data streams, it is extremely possible that one or more of these data streams, which may contain useful information, may be completely ignored, thus incurring errors that could have been avoided.

An issue which has not been touched upon at all in this paper is that of noise enhancement. Noise typically appears as fine detail in an image, and the MSR (and SSR) both will enhance it especially when small scale factors are used. Eventually, the acceptability of noise is observer and application dependent and reduces to this simple determination: Is more information *with* noise better than less information *without* noise? As long

as the noise does not contaminate the information in the signal, the authors prefer more information with noise to less information without. However the choice, as was said earlier, is user and application dependent.

Acknowledgments

Dr. Rahman's work was supported with the NASA cooperative agreement NCC-1-01030.

REFERENCES

1. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.* **83**, pp. 3078–3080, 1986.
2. E. Land, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Nat. Acad. Sci.* **80**, pp. 5163–5169, 1983.
3. E. Land, "Recent advances in retinex theory," *Vision Research* **26**(1), pp. 7–21, 1986.
4. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing* **6**, pp. 451–462, March 1997.
5. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965–976, July 1997.
6. B. D. Thompson, Z. Rahman, and S. K. Park, "A multi-scale retinex for improved performance in multi-spectral image classification," in *Visual Information Processing IX*, Proc. SPIE 4041, 2000.
7. B. D. Thompson, Z. Rahman, and S. K. Park, "Retinex preprocessing for improved multi-spectral image classification," in *Visual Information Processing VIII*, Proc. SPIE 3716, 1999.
8. C. L. Tiana, J. R. Kerr, and S. D. Harrah, "Multispectral uncooled infrared enhanced-vision system for flight test," in *Enhanced and Synthetic Vision 2001*, J. G. Verly, ed., pp. 231–236, Proc. SPIE 4363, 2001.
9. L. Brown, "A survey of image registration techniques," *ACM Computing Surveys* **24**, pp. 325–376, December 1992.
10. P. K. Kaiser and R. M. Boynton, *Human Color Vision*, Optical Society of America, Washington, DC, second ed., 1996.

Retinex processing for automatic image enhancement

Z. Rahman, D. J. Jobson, G. A. Woodell

Human Vision and Electronic Imaging VII

SPIE Symposium on Electronic Imaging

Proceedings SPIE 4662, pp. 390-401

San Jose, CA (2002)

Retinex Processing for Automatic Image Enhancement

Zia-ur Rahman[†], Daniel J. Jobson[‡], Glenn A. Woodell[‡]

[†]College of William & Mary, Department of Computer Science, Williamsburg, VA 23187.
[‡]NASA Langley Research Center, Hampton, Virginia 23681.

ABSTRACT

In the last published concept (1986) for a Retinex computation, Edwin Land introduced a center/surround spatial form, which was inspired by the receptive field structures of neurophysiology. With this as our starting point we have over the years developed this concept into a full scale automatic image enhancement algorithm—the Multi-Scale Retinex with Color Restoration (MSRCR) which combines color constancy with local contrast/lightness enhancement to transform digital images into renditions that approach the realism of direct scene observation. The MSRCR algorithm has proven to be quite general purpose, and very resilient to common forms of image pre-processing such as reasonable ranges of gamma and contrast stretch transformations. More recently we have been exploring the fundamental scientific implications of this form of image processing, namely: (i) the visual inadequacy of the linear representation of digital images, (ii) the existence of a canonical or statistical ideal visual image, and (iii) new measures of visual quality based upon these insights derived from our extensive experience with MSRCR enhanced images. The lattermost serves as the basis for future schemes for automating visual assessment—a primitive first step in bringing visual intelligence to computers.

1. INTRODUCTION

The idea of the retinex was conceived by Edwin Land[1] as a model of the lightness and color perception of human vision. Through the years, Land[2,3] evolved the concept from a random walk computation to its last form as a center/surround spatially opponent operation[3] which is related to the neurophysiological functions of individual neurons in the primate retina, lateral geniculate nucleus, and cerebral cortex. Subsequently Hurlbert[4–6] studied the properties of this form of retinex and other lightness theories and found a common mathematical foundation which possesses some excellent properties but cannot actually compute reflectance for arbitrary scenes. Certain scenes violate the "gray-world" assumption which requires that the average reflectances in the surround be equal in the three spectral color bands. For example, scenes that are dominated by one color—"monochromes"—clearly violate this assumption and are forced to be gray by the retinex computation. Hurlbert further studied the lightness problem as a learning problem for artificial neural networks and found that the solution produced was a center/ surround spatial form. This suggests the possibility that the spatial opponency of the center/surround is a general solution to estimating relative reflectances for arbitrary lighting conditions. At the same time it is equally clear that human vision does not determine relative reflectance but rather a context dependent relative reflectance since surfaces in shadow do not appear to be the same lightness as the same surface when lit. Moore et al.[7,8] took up the retinex problem as a natural implementation for analog VLSI resistive networks and found that color rendition was dependent on scene content. In each study, the consistent theoretical viewpoint was to perform all spatial processing within each spectral band and prohibit any interactions between spatial and spectral processing. This restriction provides very strong global color constancy

In our research[9–16] we do not use the retinex theory as a model for human vision color constancy. Rather, we use it as a platform for synthesizing local contrast improvement, color constancy, and lightness/color rendition as a goal for digital image enhancement. The intent is to transform the visual characteristics of the recorded digital image so that the rendition of the transformed image approaches that of the direct observation of scenes. Special emphasis is placed on increasing the local contrast in dark zones of the recorded image so that it would match our perception of wide dynamic range scenes, e.g., scenes which contain objects that are partly in sunlight

ZR: zrahman@cs.wm.edu; DJJ: d.j.jobson@larc.nasa.gov; GAW: g.a.woodell@larc.nasa.gov

and partly in shadow. Basic study of the properties of the center/surround retinex led us in the direction of using a Gaussian surround used by Hurlbert[4–6] as opposed to the $1/r^2$ surround originally proposed by Land[2,3] or the exponential surround used by Moore[7,8] for analog VLSI resistive networks. Since the width of the surround affects the rendition of the processed image, multiple scale surrounds were found to be necessary to provide a visually acceptable balance between dynamic range compression and graceful tonal rendition. This is discussed in more detail in Section 2.

The final visual defect in performance was the color "graying" due to global and regional violations of the gray-world assumption intrinsic to retinex theory. A color restoration was essential for correcting this and took the form of a log spectral operation similar to the log spatial operation of the center/surround. This produces an interaction between spatial and spectral processing and results in a tradeoff between strength of color constancy and color rendition. The color restoration yields a modest relaxation in color constancy perhaps comparable to human color vision's perceptual performance. This is discussed in more detail in Section 3.

In the course of our experiments, we noted that the commonly accepted linear representation of a scene's radiometric characteristics often fails to encompass its full dynamic range, resulting in images that either have saturated bright regions to compensate for the dark regions, or clipped dark regions in order to compensate for the bright regions. A nonlinear representation such as the MSRCR provides th e necessary dynamic range compression needed to encompass the full dynamic range of the scene. Section 4 lays out these ideas in more detail.

2. THE MULTI-SCALE RETINEX

The basic form of the Multi-scale retinex (MSR) is given by

$$R_i(x_1, x_2) = \sum_{k=1}^{K} W_k \left(\log\ I_i(x_1, x_2) - \log\left[F_k(x_1, x_2) * I_i(x_1, x_2)\right]\right) \quad i = 1, \cdots, N \tag{1}$$

where the sub-index i represents the i^{th} spectral band, N is the number of spectral bands—$N = 1$ for grayscale images and $N = 3$ for typical color images. In the latter case, $i \in R, G, B$—I is the input image, R is the output of the MSR process, F_k represents the k^{th} surround function, W_k are the weights associated with F_k, K is the number of surround functions, or scales, and $*$ represents the convolution operator. The surround functions, F_k are given as:

$$F_k(x_1, x_2) = \kappa \exp[-(x_1^2 + x_2^2)/\sigma_k^2],$$

where σ_k are the scales that control the extent of the surround—smaller values of σ_k lead to narrower surrounds—and $\kappa = 1/\left(\sum_{x_1}\sum_{x_2} F(x_1, x_2)\right)$.

As mentioned in Section 1, we found that multiple surrounds were necessary in order to achieve a graceful balance between dynamic range compression and tonal rendition. The number of scales used for the MSR is, of course, application dependent. We have found empirically, however, that a combination of three scales representing narrow, medium, and wide surrounds is sufficient to provide both dynamic range compression and tonal rendition. Figure 1 shows the input image, the output of the MSR and the outputs when the different surround functions are applied to the original image. These are obtained by setting $k = 1$ and $W_k = 1.0$ in Equation 1. As is evident from Figure 1, none of the individual scales attains the goal that we are trying to achieve: visual realism. The narrow and medium surround cases are self-explanatory. The wide-surround case deserves some discussion because it is a "nice" output image. However, the lack of dynamic range obscures the features that were visible to the observer, hence it fails the test. The MSR processed image uses features from all three scales to provide simultaneous dynamic range and tonal rendition.

3. MSR WITH COLOR RESTORATION

The general effect of retinex processing on images with regional or global gray-world violations is a "graying out" of the image either in specific regions or globally. This desaturation of color can, in some cases, be severe therefore we can consider the desired color computation as a color restoration, which should produce good color

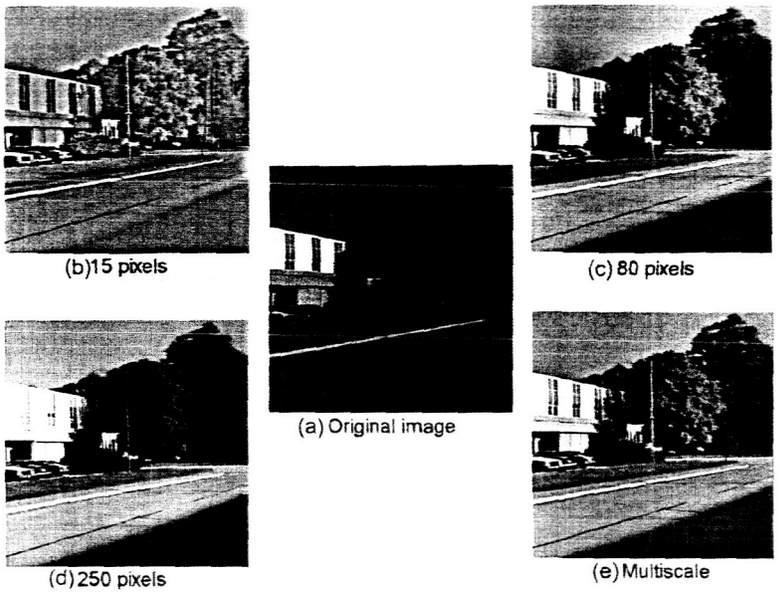


Figure 1. (a) The original input (b) Narrow surround (c) Medium surround (d) Wide surround (e) MSR output. The narrow-surround acts as a high-pass filter, capturing all the fine details in the image but at a severe loss of tonal information. The wide-surround captures all the fine tonal information but at the cost of dynamic range. The medium-surround captures both dynamic range and tonal information. The MSR is the average of the three renditions.

rendition for images with any degree of graying. In addition we would like for the correction to preserve a reasonable degree of color constancy since that is one of the basic motivations for the retinex. Color constancy is known to be imperfect in human visual perception, so some level of illuminant color dependency is acceptable provided it is much lower than the physical spectrophotometric variations. Ultimately this is a matter of image quality and color dependency is tolerable to the extent that the visual defect is not visually too strong.

We consider the foundations of colorimetry[17] even though it is often considered to be in direct opposition to color constancy models and is felt to describe only the so-called "aperture mode" of color perception, i.e. restricted to the perception of color lights rather than color surfaces.[18] The reason for this choice is simply that it serves as a foundation for creating a relative color space and in doing this uses ratios that are less dependent on illuminant spectral distributions than raw spectrophotometry. We compute a color restoration factor, α based on the following transform:

$$\alpha_i(x_1, x_2) = f\left(I_i(x_1, x_2) / \sum_{n=1}^{N} I_n(x_1, x_2)\right), \tag{2}$$

where $\alpha_i(x_1, x_2)$ is the color restoration coefficient in the i^{th} spectral band, N is the number of spectral bands, I_i is the i^{th} spectral band in the input image, and $f(\,)$ is some mapping function. In a purely empirical fashion this was tested on several retinexed images to gain a sense of the visual impact. This proved to restore color rendition, encompassing both saturated and less saturated colors. Adding this to equation 1, the Multiscale Retinex With Color Restoration (MSRCR) is given by:

$$R_i(x_1, x_2) = \alpha_i(x_1, x_2) \sum_{k=1}^{K} W_k \left(\log\ I_i(x_1, x_2) - \log\left[F_k(x_1, x_2) * I_i(x_1, x_2)\right]\right). \tag{3}$$

The results of applying this transformation to the "monochrome" images are shown in Figure 2.

While we have called this additional color computation a "restoration" we have noticed in retrospect that depending upon the form of $f(\,)$, this can be considered as a spectral analog to the spatial operation of the

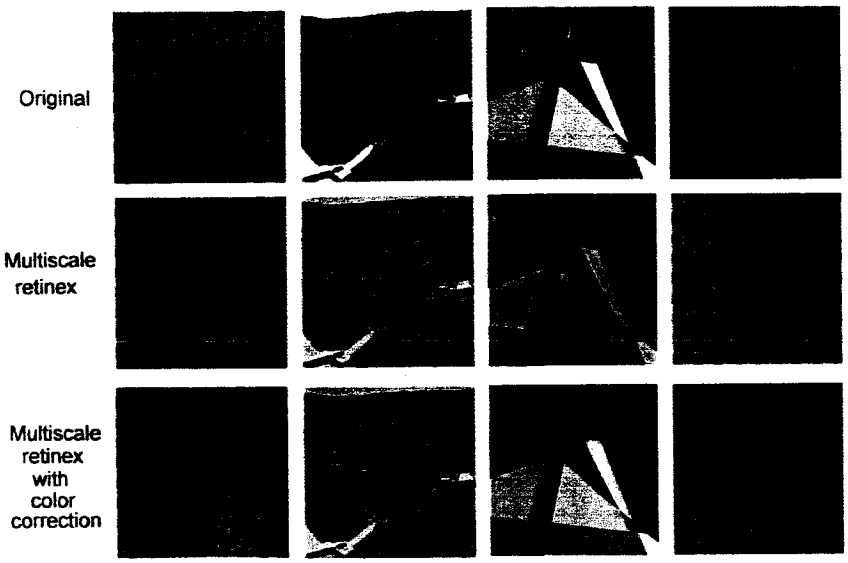


Figure 2. (Top row) Scenes that violate the gray-world assumption; (Middle row) the MSR output; note the graying of large areas of monochromes; (Bottom row) The MSRCR output; note that color constancy is diluted in order to achieve correct tonal rendition.

retinex itself. If we use

$$\alpha_i(x_1, x_2) = \log\left(I_i(x_1, x_2) / \sum_{n=1}^{N} I_n(x_1, x_2)\right),$$

then the internal form of the Retinex process and the color restoration process is essentially the same. This mathematical and philosophical symmetry is intriguing since it suggests that there may be a unifying principle at work. Both computations are contextual in nature and highly relative and nonlinear. We can venture the speculation that the visual representation of wide ranging scenes must be a compressed mesh of contextual relationships even at the stage of lightness and color representation. This sort of information representation would certainly be expected at more abstract levels of visual processing such as form information composed of edges, links, and the like but is surprising for a representation so closely related to the raw image. Perhaps in some way this front-end computation can serve later stages in a presumed hierarchy of machine vision operations that would ultimately need to be capable of such elusive goals as resilient object recognition.

4. THE MSRCR AND DIRECT VIEWING OF SCENES

Our work with the retinex[12,13] led us away from the world of color and into the world of contrast/lightness perception of complex natural scenes. While the MSRCR synthesizes color constancy, dynamic range compression, and the enhancement of contrast and lightness, the emphasis here is on the latter. We have used the MSRCR on many tens of thousands of images and find that it brings the perception of dark zones in recorded images up in local lightness and contrast to the degree needed to mimic direct scene viewing. Only images with very modest dynamic ranges do not need such enhancement and for these the exposure must be very accurate to achieve a good visual representation. Wide ranging reflectance values in a scene, and certainly, strong lighting variations demand a rather strong enhancement to achieve anything like the visual realism of direct observation. The dynamic range compression of the retinex computation is the basis for the contrast/lightness enhancement. The generic character of the computation is the basis for using it as an automatic enhancement. A few examples of retinex enhancements will serve to convey the degree to which images need to be improved and provide a demonstration that the MSRCR does, in fact, perform this task with considerable agility (Figure 3) and without human intervention. These examples highlight a major facet of retinex performance: intrinsically, the degree of automatic enhancement matches the degree of visual deficit in the original acquired image.

Figure 3. Retinex examples to illustrate that the strength of the enhancement matches the degree of visual deficit in the original image. (a) Subtle enhancements

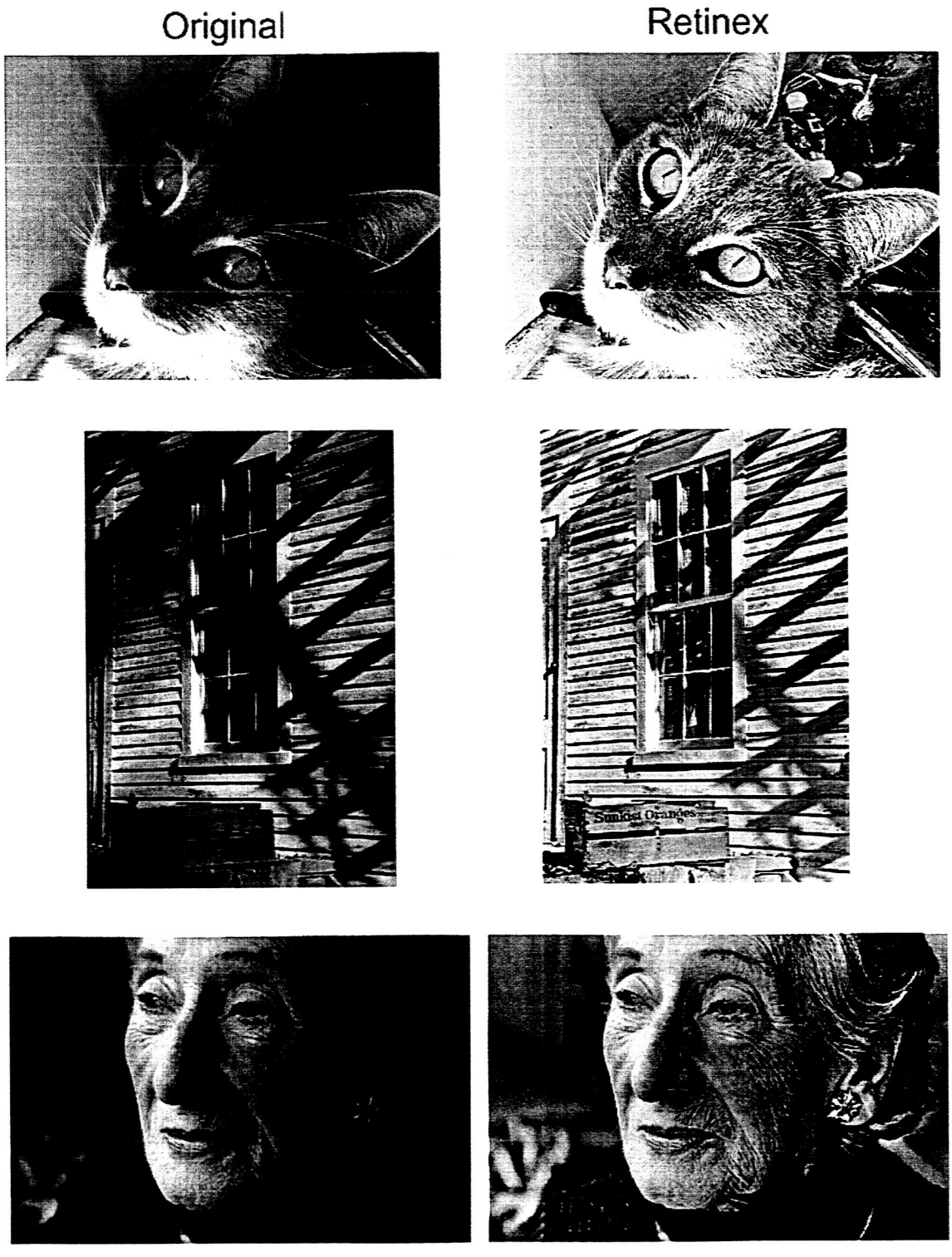


Figure 3: Continued: (b) Moderate enhancements

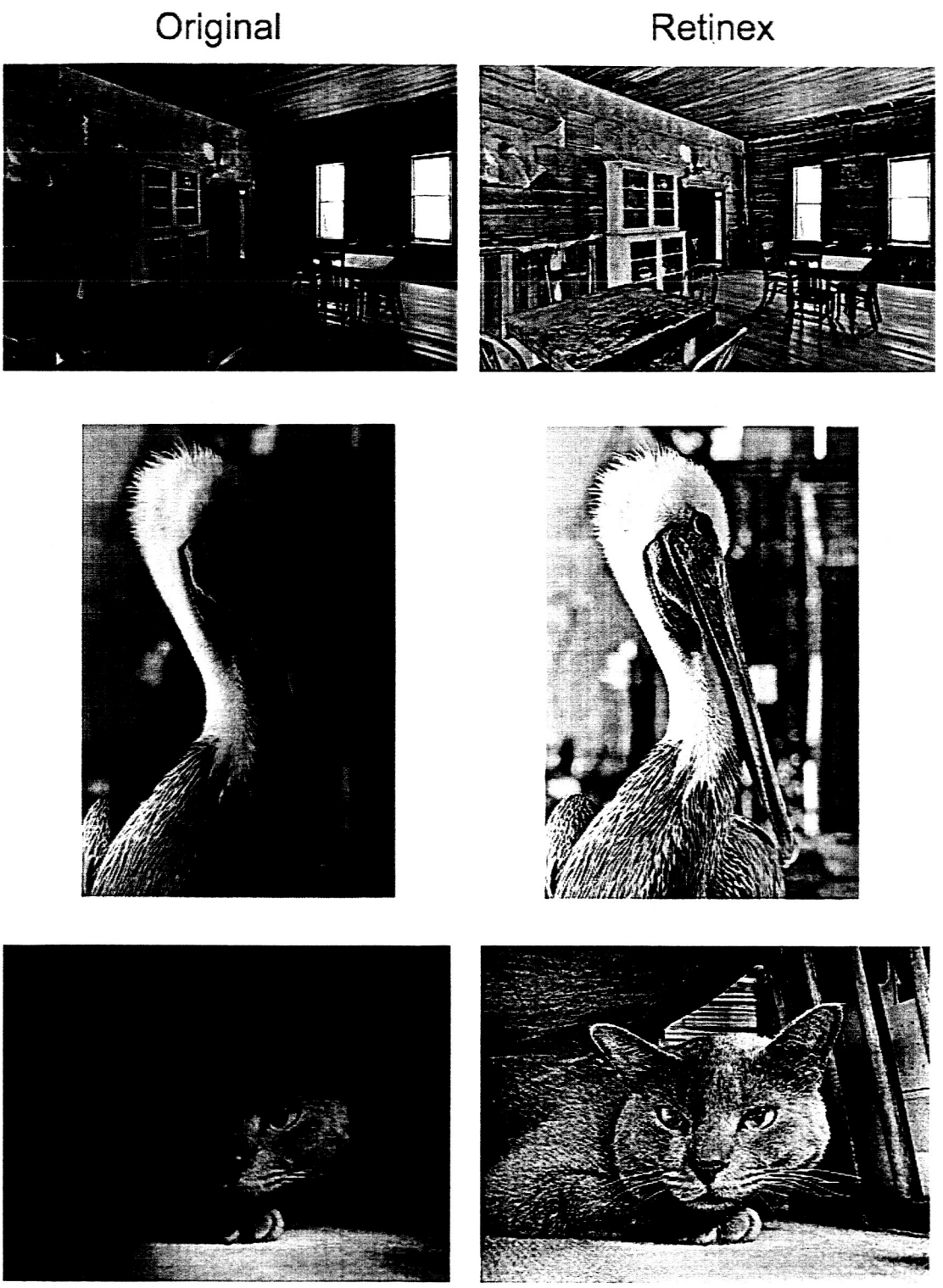


Figure 3: Continued: (c) Strong enhancements

During the course of developing this computation we were led to reexamining some of the most basic ideas about the imaging process and found that some no longer appeared tenable. Assuming that the goal of imaging is to produce a good visual representation that is comparable to the direct observation of the scene—or to provide good visibility for non-visual images such as those formed with thermal Infrared—we had to discard the idea that imaging is a replication process that produces minimal distortion of measured signals or radiometry. Instead, we had to accept the idea that *imaging is a process of profound transformation that intrinsically involves nonlinear spatial processing.* This shift arises entirely from considering the image as a visual entity and the evident visual shortcomings of the linear representation of image data (Figure 4). In general the linear representation is not a good visual representation. This is consistent with the conclusion of a study of the data handling and processing for color negative film scanning.[19] Tuijn describes the correction for all transfer functions so that the image data is linear, and then explains that this is often visually inadequate—weak in contrast and color. In order to explore this further, we displayed known linear data taken with a Nikon D1 camera in linear mode on linearized color computer monitor (gamma correction of 1.6). For a wide array of images, the displayed image is too dark (Fig 4), and the retinex enhancement (also shown for comparison) was required to produce a good visual representation. The linear representation can approach a good visual rendering for a very restricted class of scenes—those with diffuse illumination and restricted ranges of reflectances (or those where white surfaces which can be saturated). Even so, for this cooperative class a substantial degree of contrast stretching (gain/offset) is required to achieve a good visual display/print.

While image data can be quite arbitrary in a statistical sense—the histograms of images vary widely—we observe that the retinexed data were not. As noted in a previous paper,[13] histograms of MSRCR processed images tend toward a characteristic Gaussian-like shape. More recently we have studied regional means (visual lightness) and standard deviations (visual contrast) and found that they tend to converge on consistent global aggregates. This implies that a good visual representation can be associated with well-defined statistical measures for visual quality. In scientific terms, this implies the existence of a canonical visual image as a statistical practical ideal. Such a defined ideal can then serve as the basis for the automatic assessment of visual quality. While this work is still underway, we can show some preliminary results which are encouraging. By following the general idea that the retinex brings regional means and standard deviations up to higher values and that these approach an ideal goal, we have constructed tentative visual measures and performed some testing. The measures were set empirically on a small diverse test image set and then were applied to a broad array of images of all sorts. Figure 5(a) shows a sample of the automatic visual quality assessment by classification into one of three classes—poor, good, excellent. The classification scheme is based upon the map shown in Figure 5(b).

While more study and development is necessary, the early results do support the idea of a canonical visual image with well defined statistical properties. Further, the investigation indicates that the MSRCR is a valuable tool for research purposes—in this case, to define a new statistical measure of visual quality.

5. A HYPOTHETICAL DETERMINISTIC DEFINITION OF VISUAL INFORMATION

While the retinex experience provides new avenues for the study for statistical image processing, it also suggests deterministic pathways as well. The generic character of the retinex computation suggests that some new quantitative definition of visual information may be possible. A deterministic definition would contrast with previous statistical ones based upon information theory.[20,21] Specifically the MSRCR is approximately performing a log of the ratio of each pixel in each spectral band to both spatial and spectral averages. The suppression of spatial and spectral lighting variations is achieved at the expense of accepting a significant degree of context dependency. Simply put, the MSRCR appears to mimic human perception in producing color and lightness that are influenced by the visual setting in which they occur. The exchange of spatial and spectral lighting dependencies for spatio-spectral context effects appears to be a very basic element of human vision and the MSRCR computation. While we do not have a clear definition of information in a semantic sense, or visual information as some subset of all information, the idea that information is context-relationships is appealing. The additional factor of a log function suggests a compactness which may be leading in the direction of symbolic representation—the symbol being the ultimate conciseness and carrier of meaning.

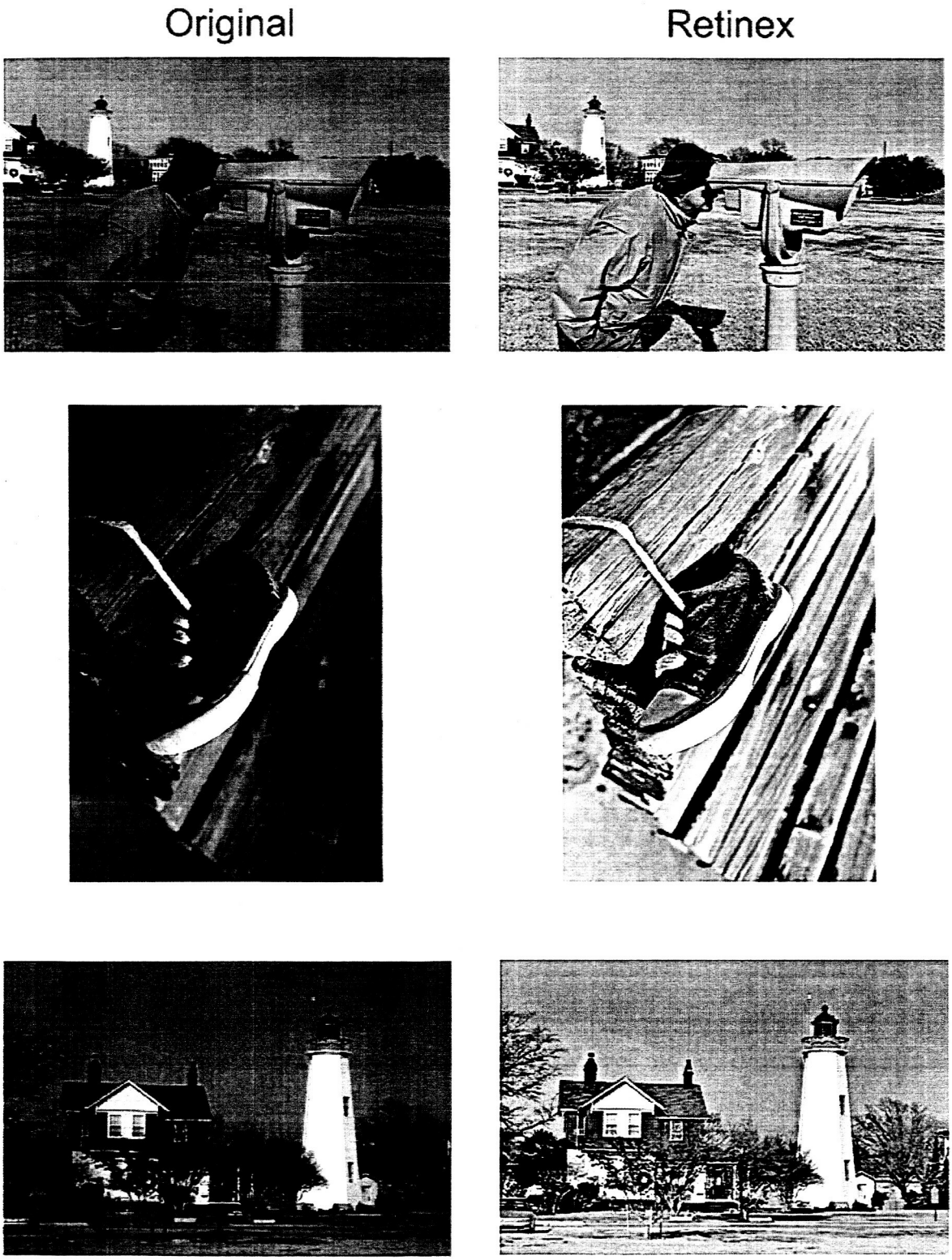


Figure 4: Visual inadequacy of the linear representation

Excellent

Good

Poor

Figure 5: Preliminary performance of Visual Measures for automating visual assessment (a) Global classes

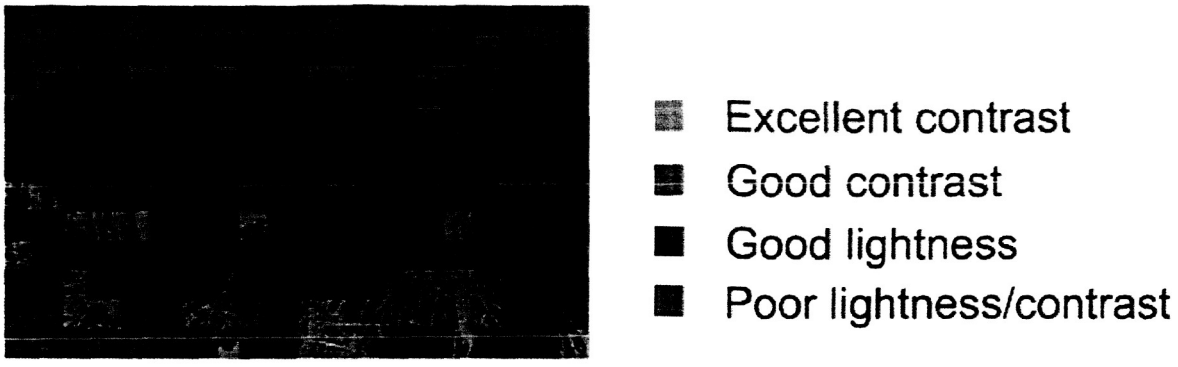


Figure 5: Continued: (b) Visual map showing regional classes

The establishment of context relationships is central to at least the senses of vision and hearing. Music seems to be based upon pitch relationships with certain ratios producing consonance or dissonance in varying degrees. Speech recognition must contend with the difficulties of speaker variations, the interdependencies of phonemes, and all manner of extraneous variations in loudness, temporal rates, degrees of clarity, and the like. For vision, the awesome task of transforming the signals of vision into the sense of vision must succeed in extracting information in the presence of all manner of extraneous variations as well as find some very concise ultimately symbolic representation. Context must be a critical element of vision information as it is in speech and music where isolated acoustical events become perceived as a fluid temporal mesh of meaningful words or melody, harmony, and rhythm. Signals are not meaningful in isolation and for vision such contextual relationships as edge connectedness, textural uniformity, and color reflectances differences seem fundamental to building a some sort of "visual information". Perhaps the retinex transformation moves one step in this direction by reducing extraneous variations, increasing spatial and spectral differences, and providing a foundation for a structure of relatedness which with subsequent processing can become symbolic.

6. CONCLUSIONS

The visual image remains an enigma full of surprises, some of which we have encountered in our experiences with retinex image processing. Though we do not understand the intricacies which allow the human vision to encompass very wide dynamic ranges, and provide color constancy, we have developed an approach that seems to mimic these behaviors. Because of this, our thinking about the imaging process has changed in basic ways:

1. Imaging should be considered as a process of transformation rather than replication with minimal distortion.
2. The statistical convergence of MSRCR image enhancements to a histogram which closely matches Gaussian distributions, leads us to postulate the existence of a canonical visual image with consistent statistical aggregate characteristics. Further, these can be used to construct entirely new visual measures which can be the basis for the automatic assessment of visual quality of arbitrary images by the computer.
3. A new deterministic definition of visual information emerges from the computational form of the retinex—namely that visual information is in some sense the log of spatial and spectral context relationships within the image.

A computation like the MSRCR appears to have two very useful properties simultaneously: a diminishment in the dependence of the appearance of the image on extraneous variables such as spatial and spectral lighting, and the construction of compact context relationships. The former is inherently useful because it can lead to better image classifications, and the latter because it shows very clearly that the appearance of a color is dependent not only on the spectral characteristics of a pixel, but also its surround. Together, these properties may be able to provide a basis for bringing more advanced levels of visual intelligence into computing.

Acknowledgments

Dr. Rahman's work was supported with the NASA co-operative agreement NCC-1-01030.

REFERENCES

1. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.* **83**, pp. 3078–3080, 1986.
2. E. Land, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Nat. Acad. Sci.* **80**, pp. 5163–5169, 1983.
3. E. Land, "Recent advances in retinex theory," *Vision Research* **26**(1), pp. 7–21, 1986.
4. A. C. Hurlbert, *The Computation of Color*. PhD thesis, Massachusetts Institute of Technology, September 1989.
5. A. C. Hurlbert, "Formal connections between lightness algorithms," *Journal of the Optical Society of America A* **3**, pp. 1684–1693, 1986.
6. A. C. Hurlbert and T. Poggio, "Synthesizing a color algorithm from examples," *Science* **239**, pp. 482–485, 1988.
7. A. Moore, J. Allman, and R. M. Goodman, "A real-time neural system for color constancy," *IEEE Transactions on Neural Networks* **2**, pp. 237–247, March 1991.
8. A. Moore, G. Fox, J. Allman, and R. M. Goodman, "A VLSI neural network for color constancy," in *Advances in Neural Information Processing 3*, D. S. Touretzky and R. Lippman, eds., pp. 370–376, Morgan Kaufmann, San Mateo, CA, 1991.
9. Z. Rahman, D. Jobson, and G. A. Woodell, "Multiscale retinex for color image enhancement," in *Proceedings of the IEEE International Conference on Image Processing*, IEEE, 1996.
10. Z. Rahman, D. Jobson, and G. A. Woodell, "Multiscale retinex for color rendition and dynamic range compression," in *Applications of Digital Image Processing XIX*, A. G. Tescher, ed., Proc. SPIE 2847, 1996.
11. D. Jobson, Z. Rahman, and G. A. Woodell, "Retinex image processing: Improved fidelity for direct visual observation," in *Proceedings of the IS&T Fourth color Imaging Conference: Color Science, Systems, and Applications*, pp. 124–126, IS&T, 1996.
12. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing* **6**, pp. 451–462, March 1997.
13. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965–976, July 1997.
14. Z. Rahman, G. A. Woodell, and D. Jobson, "A comparison of the multiscale retinex with other image enhancement techniques," in *Proceedings of the IS&T 50th Anniversary Conference*, pp. 426–431, IS&T, 1997.
15. Z. Rahman, D. Jobson, and G. A. Woodell, "Resliency of the multiscale retinex image enhancement algorithm," in *Proceedings of the IS&T Sixth color Imaging Conference: Color Science, Systems, and Applications*, pp. 129–134, IS&T, 1998.
16. D. Jobson, Z. Rahman, and G. A. Woodell, "Spatial aspect of color and scientific implications of retinex image processing," in *Visual Information Processing X*, S. K. Park, Z. Rahman, and R. A. Schowengerdt, eds., pp. 117–128, Proc. SPIE 4388, 2001.
17. W. D. Wright, *The Measurement of Colour*, Hilger and Watts, London, second ed., 1958.
18. P. Lennie and M. D. D'Zmura, "Mechanisms of color vision," in *CRC Critical Reviews of Neurobiology*, vol. 3, pp. 333–400, 1988.
19. C. Tuijn, "Scanning color negatives," in *Proceedings of the IS&T Fourth color Imaging Conference: Color Science, Systems, and Applications*, pp. 33–38, IS&T, 1996.
20. F. O. Huck, C. L. Fales, and Z. Rahman, "Information theory of visual communication," *Philosophical Transactions of the Royal Society of London A* **354**, pp. 2193–2248, Oct. 1996.
21. F. O. Huck, C. L. Fales, and Z. Rahman, *Visual Communication: An Information Theory Approach*, Kluwer Academic Publishers, Boston, MA, 1997.

The Spatial Aspect of Color and Scientific Implications of Retinex Image Processing

D. J. Jobson, Z. Rahman, and G. A. Woodell

SPIE International Symposium on AeroSense

Proceedings SPIE 4388, pp. 118–128

Orlando, FL (2001)

The Spatial Aspect of Color and Scientific Implications of Retinex Image Processing

Daniel J. Jobson[†], Zia-ur Rahman[‡], Glenn A. Woodell[†]

[†]NASA Langley Research Center, Hampton, Virginia 23681.
[‡]College of William & Mary, Department of Computer Science, Williamsburg, Virginia 23187.

ABSTRACT

The history of the spatial aspect of color perception is reviewed in order to lay a foundation for the discussion of retinex image processing. While retinex computations were originally conceived as a model for color constancy in human vision, the impact on local contrast and lightness is even more pronounced than the compensation for changes in the spectral distribution of scene illuminants. In the multiscale retinex with color restoration (MSRCR), the goal of the computation is fidelity to the direct observation of scenes. The primary visual shortcoming of the recorded image is that dark zones such as shadow zones are perceived with much lower contrast and lightness than for the direct viewing of scenes. Extensive development and testing of the MSRCR led us to form several hypotheses about imaging which appear to be basic and general in nature. These are: (1) the linear representation of the image is not usually a good visual representation, (2) retinex image enhancements tend to approach a statistical ideal which suggests the existence of a canonical "visual image", and (3) the mathematical form of the MSRCR suggests a deterministic definition of visual information which is the log of the spectral and spatial context ratios for any given image. These ideas imply that the imaging process should be thought of, not as a replication process whose goal is minimal distortion, but rather as a profound non-linear transformation process whose goal is a statistical ideal visual representation. These insights suggest new directions for practical advances in bringing higher levels of visual intelligence to the world of computing.

1. INTRODUCTION—THE SPATIAL ASPECT OF COLOR

Visual perception is replete with surprises. Among the most basic is that color perception has a strong spatial dependency rather than being purely spectral in nature. This was demonstrated dramatically by Edwin Land with color constancy experiments. By manipulating the color of light sources so that green and red reflectance patches were spectrally identical, he showed that they are still seen as green and red. In fact one would be hard pressed to define any explanation of human vision's color constancy which does not demand an interaction between spectral and spatial processing of images.

Land postulated a number of variations[1–4] of his retinex theory which culminated in a last version[3] taking the form of a non-linear center/surround. This form was the starting point for our MSRCR which elaborates the concept to include multiple scales of surrounds and defines a color restoration which overcomes the fundamental practical limitation of the gray-world assumption intrinsic to the retinex concept. Multiple scales and the color restoration were required to produce an general purpose and automatic image enhancement method with graceful tonal and color performance.

More recently, a striking new observation has arisen from a study of the color perception of the optical spectra. Smeulders et al.[5] noted discrepancies in the historical data of various key figures in color physics—Newton, Young, and Helmholtz. These all centered of differing descriptions of the colors seen in spectra projected from prisms and were traced to differences in the manner of demarcation used. A thorough

DJJ: d.j.jobson@larc.nasa.gov; co-authors: ZR: zrahman@cs.wm.edu; GAW: g.a.woodell@larc.nasa.gov

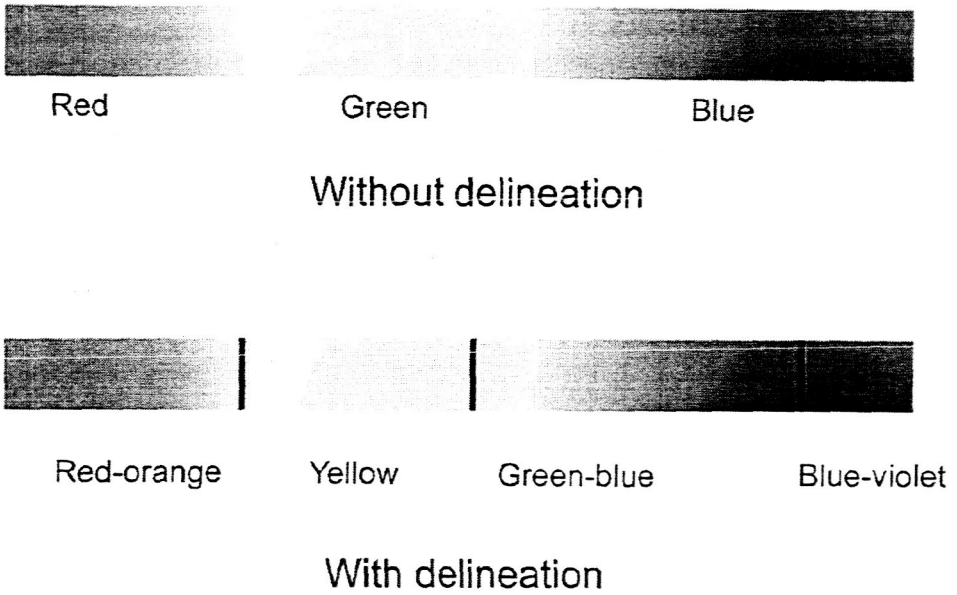


Figure 1. Perception of spectral color

study of the perception of spectral color led to the conclusion that only three primary colors are seen in a spectrum without delineation (Figure 1). When lines are added to the spectrum, the full color gamut begins to emerge- more lines, more colors up to a saturation point where the addition of more lines brings about fewer and fewer additional hues. The implication like that of Land's work is that there is a strong influence of spatial processing in color perception. In this case, the conclusion is very basic—that spatial structure is essential for full color perception.

2. RETINEX IMAGE PROCESSING—CONTRAST AND LIGHTNESS FIDELITY TO DIRECT VIEWING OF SCENES

Our own work with retinex image processing[6,7] leads us away from the world of color and into the world of contrast/lightness perception of complex natural scenes. While the MSRCR synthesizes color constancy, dynamic range compression, and the enhancement of contrast and lightness—the emphasis here is on the latter. We have used the MSRCR with many tens of thousands of test images and find it to be a generic image enhancement computation. Specifically, it does bring the perception of dark zones in recorded images up in local lightness and contrast to the degree needed to mimic direct scene viewing. Only images with very modest dynamic ranges do not need enhancement and for these the exposure must be very accurate to achieve a good visual representation. Even wide ranging reflectance values in a scene, and certainly strong lighting variations demand a rather strong enhancement to achieve any thing like the visual realism of direct observation. The dynamic range compression of the retinex computation is the basis for the contrast/lightness enhancement. The generic character of the computation is the basis for using it as an automatic enhancement. A few examples of retinex enhancements will serve to convey the degree to which images need to be improves and provide a demonstration that the MSRCR does, in fact, perform this task with considerable agility (Figure 2) and without human intervention. These examples highlight a major facet of retinex performance: intrinsically, the degree of automatic enhancement matches the degree of visual deficit in the original acquired image.

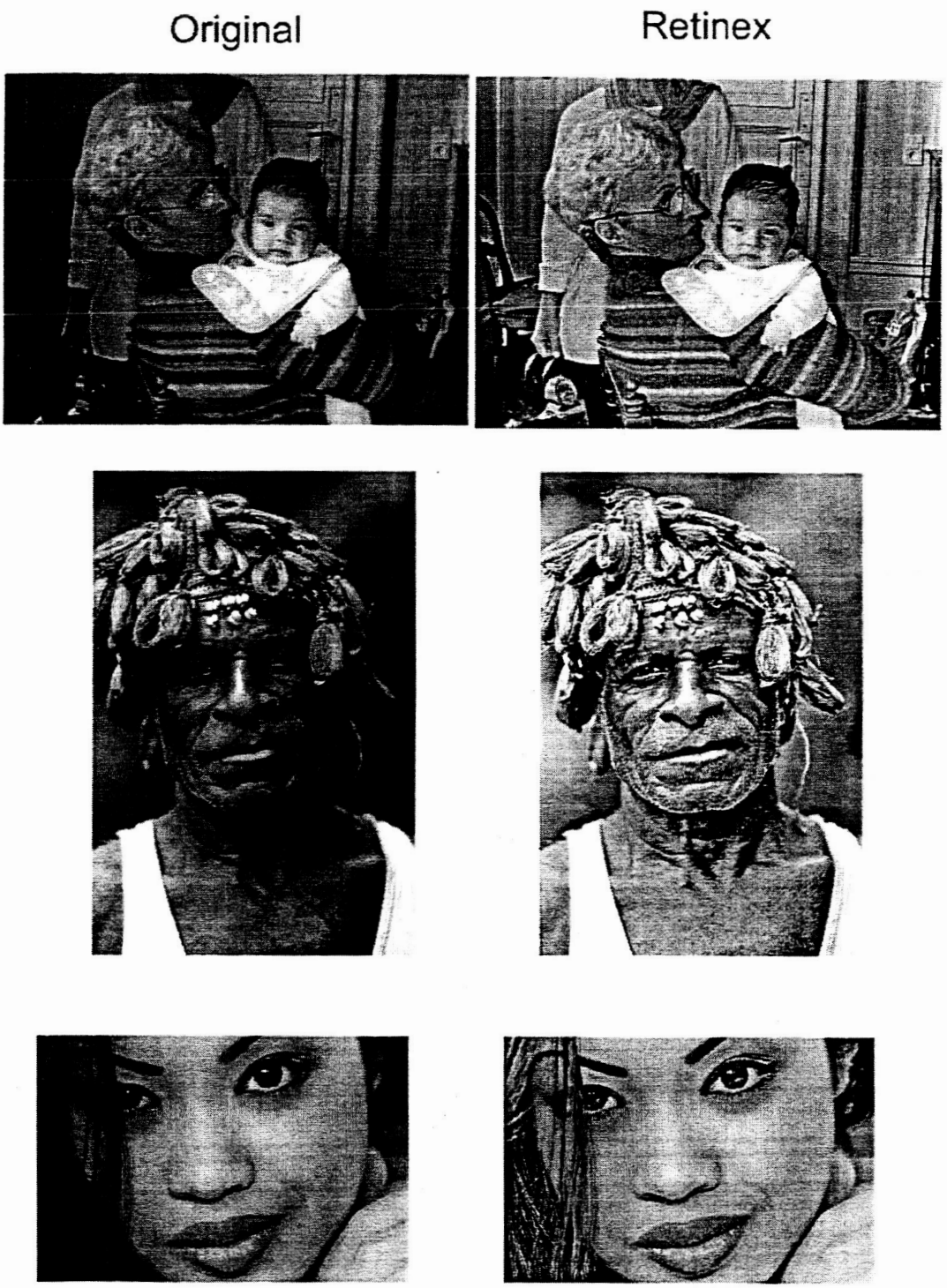


Figure 2. Retinex examples to illustrate that the strength of the enhancement matches the degree of visual deficit in the original image. (a) Subtle enhancements

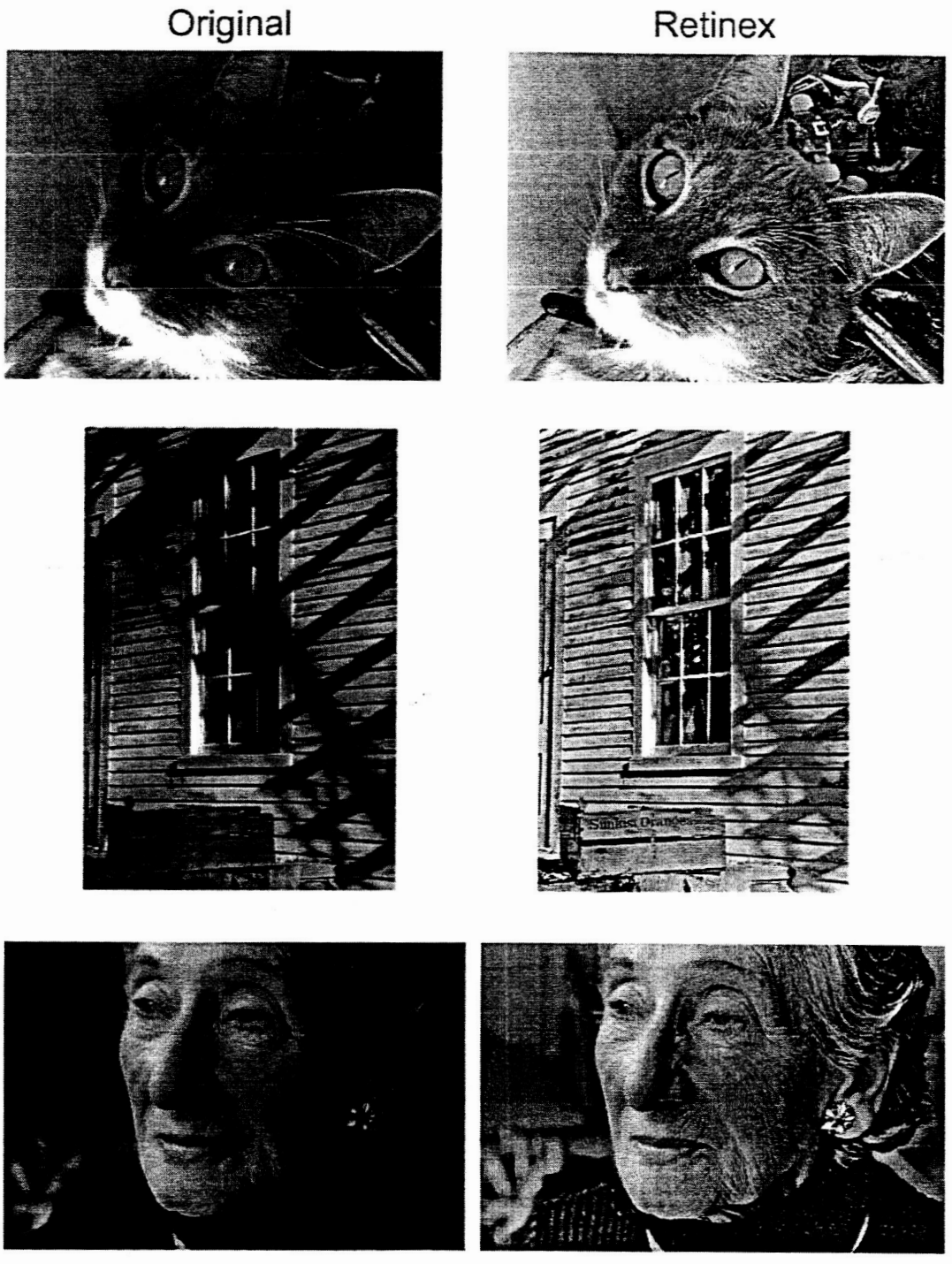


Figure 2. Continued: (b) Moderate enhancements

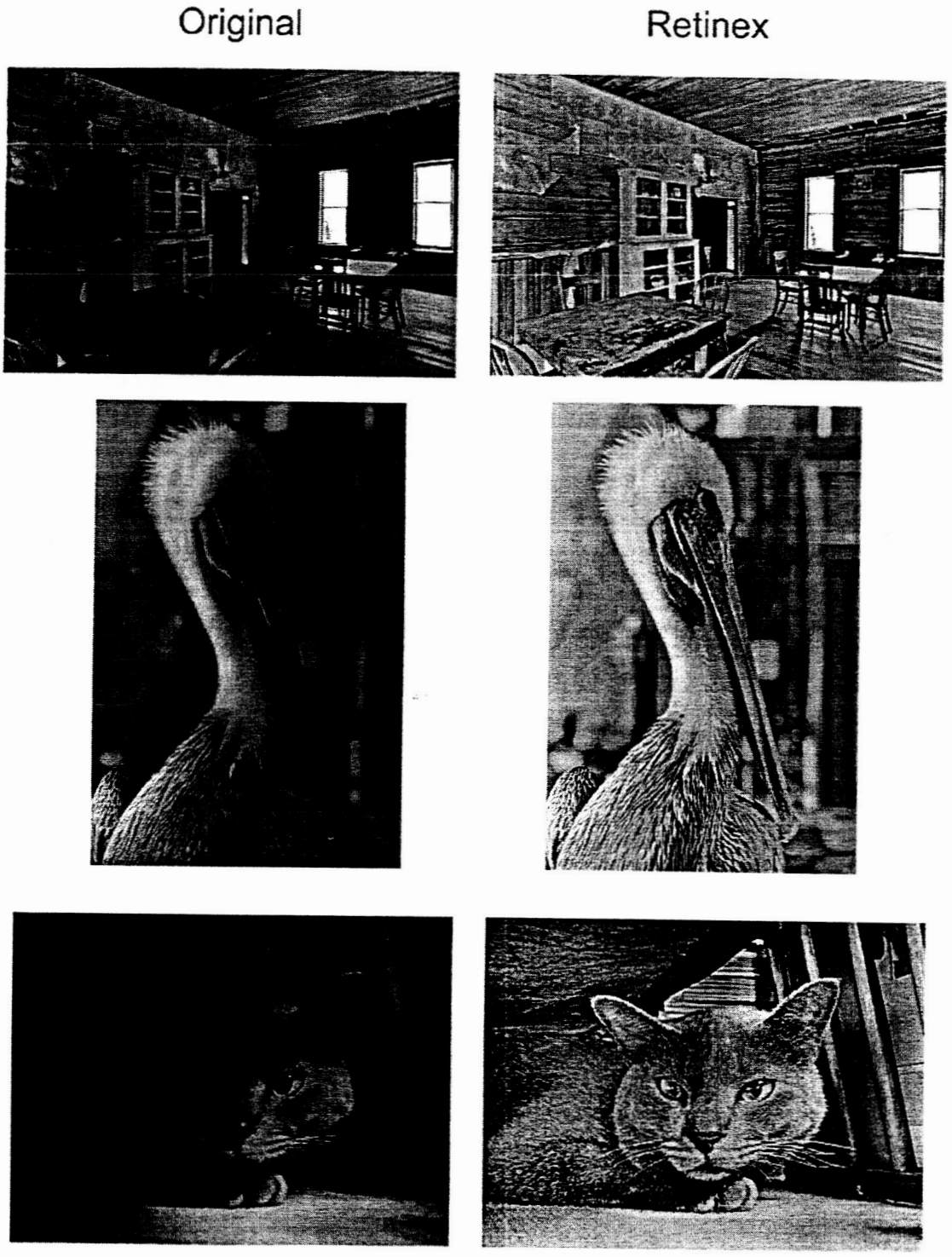


Figure 2. Continued: (b) Stong enhancements

The design of the MSRCR was defined by experiments whereby computational parameters were adjusted until the processed image display or print compared well with the direct viewing of test scenes. This was then tested successfully on a large very diverse battery of several thousand test images in order to ferret out any quirks in performance or rare pathologies. Since most digital images do not come with any knowledge of pre-processing which may have been applied at the camera or scanner software driver level or by subsequent image processing tools, the impact of retinex performance on various commonplace forms of preprocessing was studied. While distortions were observed, they were generally mild and the MSRCR was reasonably resilient to a range of gamma and contrast stretch operations applied prior to the retinex computation.[8] Ideally the retinex should be applied to linear data with only dark offsets removed and with no image compression or high quality levels of JPEG. The impact of preprocessing (either gamma or contrast stretch) is to shift the optimum post-retinex gain/offset values. Should preprocessing be consistent, then a resetting of the post-retinex gain/offset can be used to produce best performance on preprocessed image data.

While the MSRCR was developed with general purpose color imaging in mind, it solves such a fundamental problem of imaging—good visibility across a wide dynamic range of data—that it is useful for scientific and other special purpose imaging applications such as medical, forensic, surveillance, and remote sensing applications (Figure 3). It has been shown in previous studies[9,10] to improve multispectral classification accuracies when used as as a front-end to supervised training classification schemes where extensive ground truth was available. The practical value of retinex image processing is accompanied by the emergence of new scientific insights into the visual image and the imaging process. These will be outlined in the remainder of this text.

3. SCIENTIFIC IMPLICATIONS OF THE RETINEX EXPERIENCE

3.1. The visual inadequacy of the linear representation

During the course of developing this retinex computation and testing it experimentally on large numbers of diverse images, we were forced to re-examine some of our most basic ideas about the imaging process and found that some were no longer tenable. If we assume that the goal of imaging is a good visual representation which compares to the direct observation of scenes for color imaging or simply provides good visibility for imaging outside visual spectral range (IR, MRI, Xray,etc.), we had to discard the idea that imaging is a replication process whose goal is minimal distortion of measured signals or radiometry. In place of this long-standing tradition of much of signal and image processing, we had to move to the idea that imaging is a process of profound transformation which intrinsically involves non-linear spatial processing. This shift arises entirely from considering the image as a visual entity and the evident visual shortcomings of the linear representation of image data (Figure 4). In general the linear representation is not a good visual representation. This is consistent with the conclusion of a study of the data handling and processing for color negative film scanning.[11] Tuijn describes the correction for all transfer functions so that the image data is linear, and then explains that this is often visually inadequate—weak in contrast and color. In order to explore this further, we displayed known linear data taken with a Nikon D1 camera in linear mode on linearized color computer monitor (gamma correction of 1.6). For a wide array of images, the displayed image is too dark (Fig 4), and the retinex enhancement (also shown for comparison) was required to produce a good visual representation. The linear representation can approach a good visual rendering for a very restricted class of scenes—those with diffuse illumination and restricted ranges of reflectances (or those where white surfaces which can be saturated). Even so, for this cooperative class a substantial degree of contrast stretching (gain/offset) is required to achieve a good visual display/print.

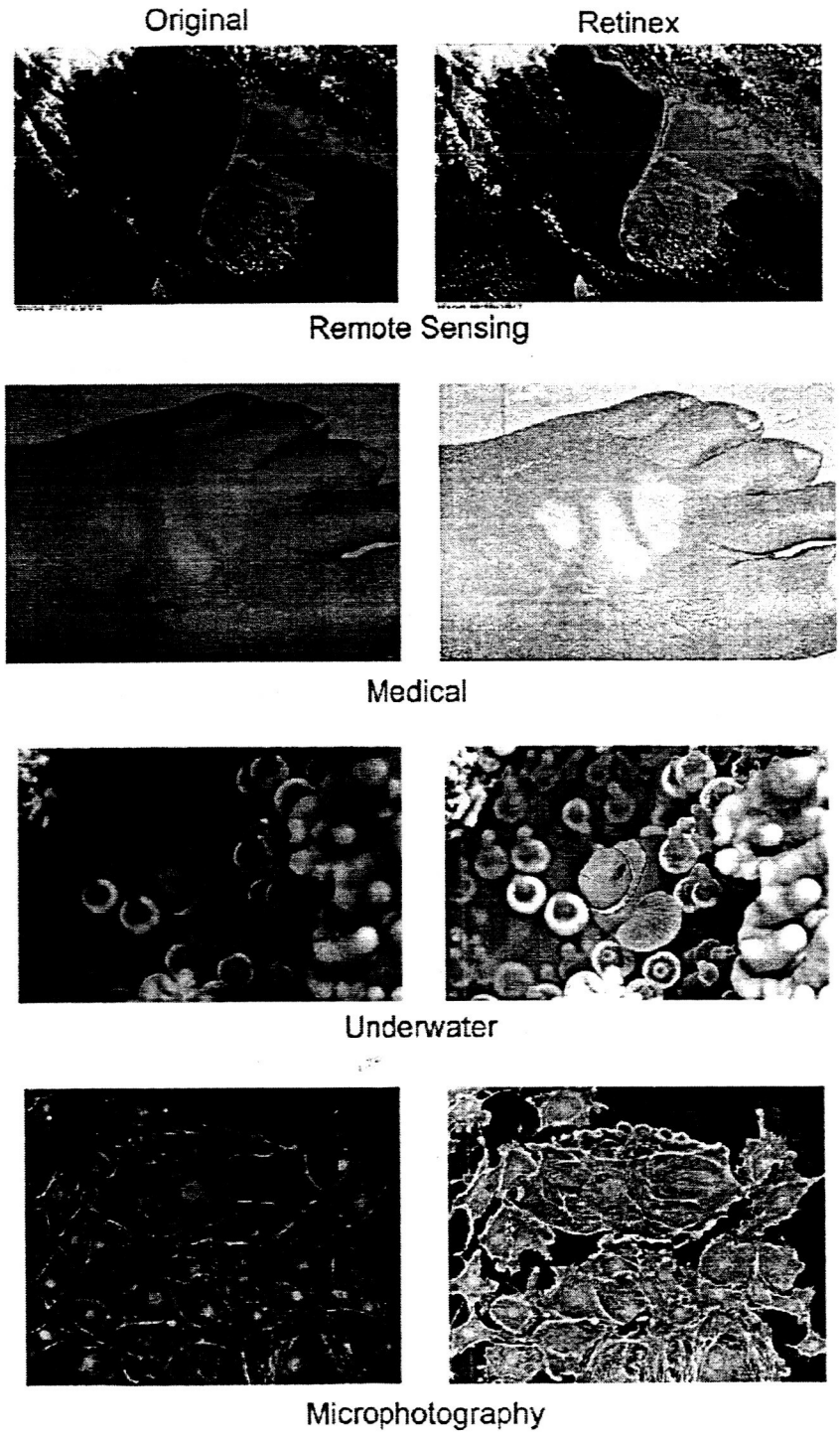


Figure 3. Illustration of specialized applications

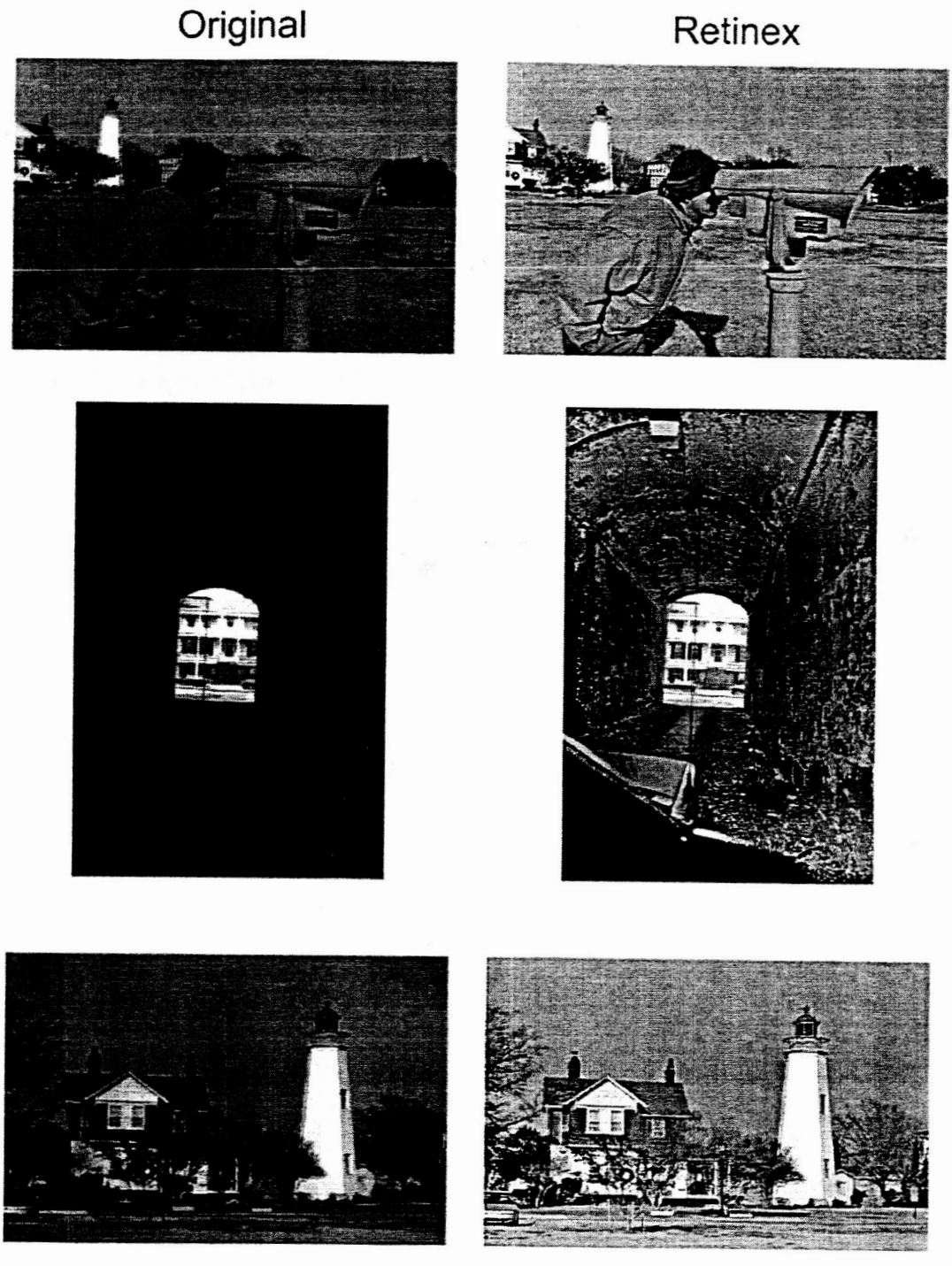


Figure 4. Visual inadequacy of the linear representation

3.2. The existence of a canonical visual image and definition of visual measures for automating visual assessment

While image data is quite arbitrary in a statistical sense, we observe that retinexed data were not. As noted in a previous paper,[7] retinex histograms tend toward a characteristic Gaussian-like shape. More recently we have studied regional means (visual lightness) and standard deviations (visual contrast) and found that they tend to converge on consistent global aggregates. This implies that a good visual representation can be associated with well-defined statistical measures for visual quality. In scientific terms, this implies the existence of a canonical visual image as a statistical practical ideal. Such a defined ideal can then serve as the basis for the automatic assessment of visual quality. While this work is still underway, we can show some preliminary results which are encouraging. By following the general idea that the retinex brings regional means and standard deviations up to higher values and that these approach an ideal goal, we have constructed tentative visual measures and performed some testing. The measures were set empirically on a small diverse test image set and then were applied to a broad array of images of all sorts. Figure 5(a) shows a sample of the automatic visual quality assessment by classification into one of three classes—poor, good, excellent. The classification scheme is based upon the map shown in Figure 5(b).

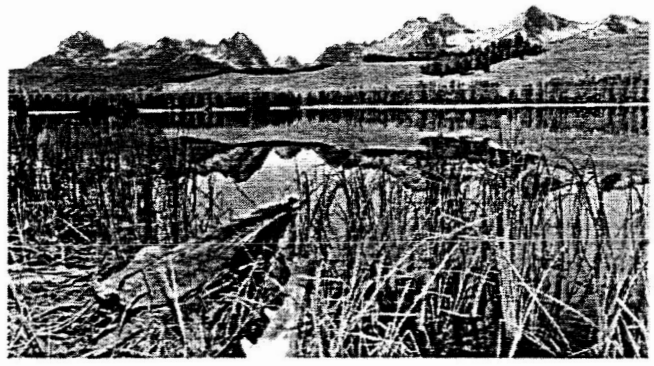
While more study and development is necessary, the early results do support the idea of a canonical visual image with well defined statistical properties. Further, the investigation indicates that the MSRCR is a valuable tool for research purposes—in this case, to define a new statistical measure of visual quality.

Currently, computers have essentially no visual intelligence. Visual measures which can enable the automatic assessment of digital images are a first step in visual intelligence. Such measures would allow the computer to determine visual quality and automate image processing at a higher level and in a more sophisticated manner than is now possible.

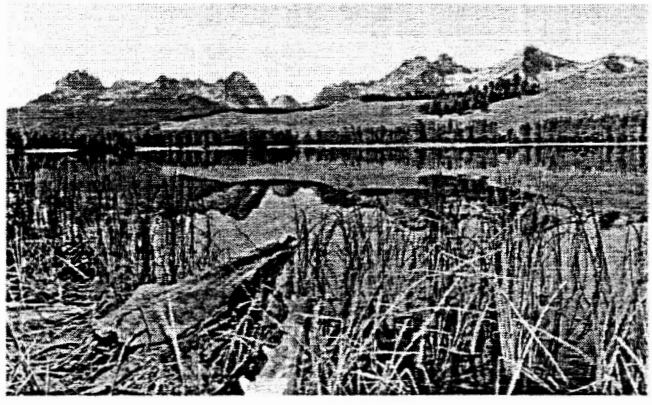
3.3. A Hypothetical Deterministic Definition of Visual Information

While the retinex experience provides new avenues for the study for statistical image processing, it also suggests deterministic pathways as well. The generic character of the retinex computation suggests that some new quantitative definition of visual information may be possible. A deterministic definition would contrast with previous statistical ones based upon information theory.[12,13] Specifically the retinex is approximately performing a log of the ratio of each pixel in each spectral band to both spatial and spectral averages. The suppression of spatial and spectral lighting variations is achieved at the expense of accepting a significant degree of context dependency. Simply put the retinex mimics human perception in producing color and lightness which which are influenced by the visual setting in which they occur. The exchange of spatial and spectral lighting dependencies for spatio-spectral context effects appears to be a very basic element of human vision and the retinex computation. While we do not have a clear definition of information in a semantic sense, or visual information as some subset of all information, the idea that information is context relationships is appealing. The additional factor of a log function suggests a compactness which may be leading in the direction of symbolic representation—the symbol being the ultimate conciseness and carrier of meaning.

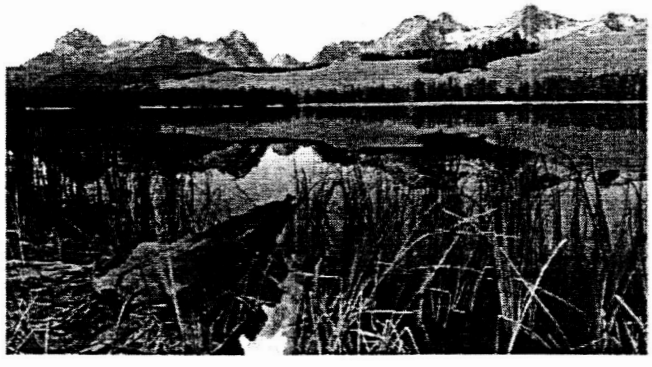
The establishment of context relationships is central to at least the senses of vision and hearing. Music seems to be based upon pitch relationships with certain ratios producing consonance or dissonance in varying degrees. Speech recognition must contend with the difficulties of speaker variations, the interdependencies of phonemes, and all manner of extraneous variations in loudness, temporal rates, degrees of clarity, and the like. For vision, the awesome task of transforming the signals of vision into the sense of vision must succeed in extracting information in the presence of all manner of extraneous variations as well as find some very concise ultimately symbolic representation. Context must be a critical element of vision information as it is in speech and music where isolated acoustical events become perceived as a



Excellent



Good



Poor

Figure 5. Preliminary performance of Visual Measures for automating visual assessment (a) Global classes

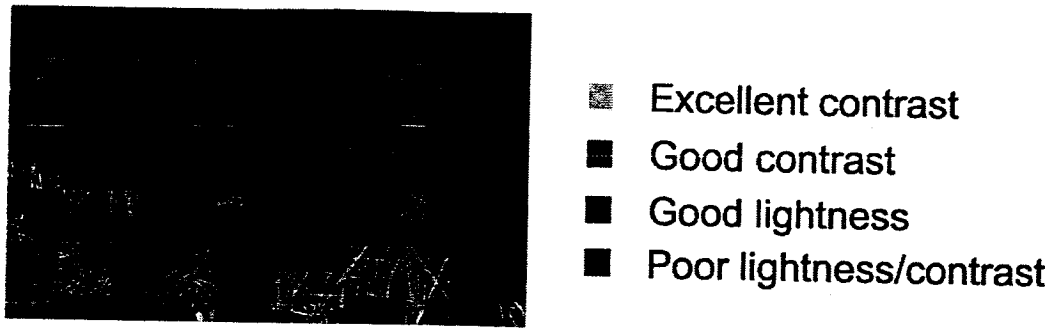


Figure 5. Continued: (b) Visual map showing regional classes

fluid temporal mesh of meaningful words or melody, harmony, and rhythm. Signals are not meaningful in isolation and for vision such contextual relationships as edge connectedness, textural uniformity, and color reflectances differences seem fundamental to building a some sort of "visual information". Perhaps the retinex transformation moves one step in this direction by reducing extraneous variations, increasing spatial and spectral differences, and providing a foundation for a structure of relatedness which with subsequent processing can become symbolic.

3.4. Impact of Retinex Processing on Sensor Design and End-to-End Processing

How should the retinex computation fit within the system of acquiring, processing, and displaying images? How does it exert an influence, if any, on sensor design and other image processing? With respect to sensor design the retinex from our practical experience, as well as from the consideration of scene radiometry, impacts the sensor design by asking for wide dynamic range image data with high signal-to-noise ratios (SNR). In order to minimize the visual distraction of emphatic noise in the retinex enhancement, wide dynamic range scenes should be acquired with high sensitivity sensors. For visible color imaging, image data dynamic ranges (at equivalent SNR) should (ideally) be in the 10–12bits range in order to encompass the scene radiometry of many everyday scenes and avoid noise visibility in the retinex enhancement of images with large very dark regions.

Ideally the retinex computation should be performed immediately after image acquisition and prior to other processing, especially data compression. Less ideally, the retinex can be applied with good results after compression if high quality levels (low compression ratios) of JPEG, for example, are used. There is, however a basic tension between the retinex computation and JPEG, in that JPEG hides artifacts in imperceivable dark zones which the retinex enhances into visibility when higher compression ratios are selected.

If the retinex is applied prior to compression by being embedded in a sensor chip or is applied to uncompressed data, the enhancement makes full use of the sensor performance envelope to achieve results that are limited only by the sensor performance itself—primarily SNR.

A front-end retinex has the advantage of bringing out the visual quality so that subsequent compression is done on this improved starting point. This reduces the likelihood that significant visual features will be distorted by JPEG or other coding artifacts.

4. CONCLUSION

The visual image remains an enigma full of surprises, some of which we encountered in our experience with retinex image processing. Our thinking about the imaging process has been changed in basic ways outlined here. The new directions stimulated are summarized as:

1. Imaging should be considered as a process of transformation rather than replication with minimal distortion. This idea arose from the visual inadequacy of the linear representation of image data.
2. The statistical convergence of retinex image enhancements led us to postulate the existence of a canonical visual image with consistent statistical aggregate characteristics. Further these can be used to construct entirely new visual measures which can be the basis for the automatic assessment of visual quality of arbitrary images by the computer.
3. A new deterministic definition of visual information emerges from the computational form of the retinex—namely that visual information is in some sense the log of spatial and spectral context relationships within the image.
4. The retinex should be applied ideally as a front-end computation prior to any data compression and achieves its fullest visual quality when sensor SNR and digitization cover the 10–12bits dynamic range required by the radiometric character of commonplace scenes.

A computation like the retinex appears to solve two problems simultaneously—the diminishing of extraneous variables such as spatial and spectral lighting and the construction of compact context relationships which may provide a basis for bringing more advanced levels of visual intelligence into computing.

REFERENCES

1. E. Land, "Color vision and the natural image," *Proc. Nat. Acad. Sci.* **45**, pp. 115–129, 1959.
2. E. Land, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Nat. Acad. Sci.* **80**, pp. 5163–5169, 1983.
3. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.* **83**, pp. 3078–3080, 1986.
4. E. Land, "Recent advances in retinex theory," *Vision Research* **26**(1), pp. 7–21, 1986.
5. N. Smeulders, F. W. Campbell, and P. R. Andrews, "The role of delienation and spatial frequency in the perceptoin of the colours of the spectrum," *Vision Research* **34**(7), pp. 927–936, 1994.
6. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing: Special Issue on Color Processing* **6**, pp. 451–462, March 1997.
7. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965–976, July 1997.
8. Z. Rahman, D. Jobson, and G. A. Woodell, "Resliency of the multiscale retinex image enhancement algorithm," in *Proceedings of the IS&T Sixth color Imaging Conference: Color Science, Systems, and Applications*, pp. 129–134, IS&T, 1998.
9. B. D. Thompson, Z. Rahman, and S. K. Park, "A multi-scale retinex for improved performance in multi-spectral image classification," in *Visual Information Processing IX*, Proc. SPIE 4041, 2000.
10. B. D. Thompson, Z. Rahman, and S. K. Park, "Retinex preprocessing for improved multi-spectral image classification," in *Visual Information Processing VIII*, Proc. SPIE 3716, 1999.
11. C. Tuijn, "Scanning color negatives," in *Proceedings of the IS&T Fourth color Imaging Conference: Color Science, Systems, and Applications*, pp. 33–38, IS&T, 1996.
12. F. O. Huck, C. L. Fales, and Z. Rahman, "Information theory of visual communication," *Philosophical Transactions of the Royal Society of London A* **354**, pp. 2193–2248, Oct. 1996.
13. F. O. Huck, C. L. Fales, and Z. Rahman, *Visual Communication: An Information Theory Approach*, Kluwer Academic Publishers, Boston, MA, 1997.